AD_____

AWARD NUMBER: DAMD17-02-1-0214

TITLE: Digital Mammography: Development of an Advanced Computer-Aided Diagnosis System for Breast Cancer Detection

PRINCIPAL INVESTIGATOR: Heang-Ping Chan, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: May 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-05-2006	Final	1 May 2002 – 30 Apr 2006

4. TITLE AND SUBTITLE

Digital Mammography: Development of an Advanced Computer-Aided Diagnosis System for Breast Cancer Detection

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-02-1-0214

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Heang-Ping Chan, Ph.D.

E-Mail: chanhp@umich.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Michigan
Ann Arbor, Michigan 48109-1274

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of the project is to develop computer-aided diagnosis (CAD) methods and systems for mammography using advanced computer vision techniques and image information fusion from multiple mammograms to improve lesion detection and characterization. When fully developed, the CAD system can assist radiologists in mammographic interpretation. During this project year, we have performed the following tasks: (1) collected databases of digital mammograms (DMs) and digitized film mammograms (DFMs) for development of the CAD systems, (2) compared the performance of the microcalcification detection system on DMs and DFMs, (3) developed two-view information fusion techniques for the mass detection system, (4) developed bilateral analysis technique for the mass detection system, and (5) evaluated the effects of reconstruction image quality on mass detection for digital tomosynthesis mammography (DTM), which is a promising new imaging modality for improving breast cancer detection. In summary, we have investigated a number of areas in CAD of mammographic lesions and evaluated new techniques for both DMs and DFMs. We have found that our new computer-vision techniques and multiple-image analysis approach can improve the performance of the CAD systems. We have also explored CAD methods for DTM and evaluated their dependence on reconstruction techniques. We will continue the development of the CAD systems for mammography in the coming years.

15. SUBJECT TERMS
breast cancer digital mammography, computer-aided diagnosis, breast cancer diagnosis

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	62	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

(3) Table of Contents

(1) Front Cover 1

(2) Standard Form (SF) 298, REPORT DOCUMENTATION PAGE 2

(3) Table of Contents 3

(4) Introduction 4

(5) Body 5

- (A) Collection of databases of digital mammograms and digitized film mammograms
- (B) CAD system for microcalcification detection on digital mammograms – comparison of detection accuracy on digitized film mammograms and digital mammograms
- (C) CAD system for mass detection on mammograms – Two view information fusion
- (D) CAD system for mass detection on mammograms – Bilateral analysis for false positive reduction
- (E) Computer-aided mass detection on digital tomosynthesis mammograms (DTM) – Dependence on reconstruction image quality

(6) Key Research Accomplishments 12

(7) Reportable Outcomes 13

(8) Conclusions 15

(9) References 16

(10) Appendix 16

(4) Introduction

Computer-aided diagnosis (CAD) has been shown to be useful as a second opinion to radiologists for breast cancer detection on mammograms. All current CAD systems have been developed for digitized screen-film mammograms (DFM). With the recent advent of full field digital mammography (FFDM) systems, it is important to develop CAD systems specifically designed for direct digital mammograms (DMs) in order to fully exploit the advantages of FFDM. Although many computer vision techniques developed for digitized films may be used for DMs, proper adaptation and extensive training of the current algorithms for the new type of images will be required. More importantly, new techniques still need to be developed to further improve the current algorithms for DFMs as well as for adapting to FFDM.

The goal of the proposed research is to develop a CAD system for breast cancer diagnosis using advanced computer vision techniques. The proposed CAD system will assist radiologists with detection and classification of breast lesions. Previous CAD methods for lesion detection and characterization are generally based on image features extracted from a single view. Our proposed approach is based on two steps: the first step uses single view detection to identify lesion candidates on individual mammograms, the second step is to fuse image information from multiple views to reduce false positives and thus to improve the overall accuracy. Although the main goal of this project is to develop a CAD system for DMs, we plan to extend the CAD development to DFMs for the following reasons: (1) digital mammography only became available in the last few years, multiple-view film mammograms with breast lesions are more commonly available in existing patient files, and (2) screen-film mammography will still be the main modality for breast cancer screening in the near future. Therefore, we will first develop the multiple-view correlation techniques for the CAD system of the DFMs. These new techniques will then be adapted to the CAD system for DMs. We believe that this approach is more efficient and we will obtain a CAD system for DMs as well as improve the CAD system for DFMs.

The following specific aims will be addressed: (1) Collection of databases of both DMs and DFMs and design of a database management system. (2) Improvement of single-view computer vision techniques for mass detection and classification in DFMs. (3) Improvement of single-view computer vision techniques for microcalcification detection and classification in DFMs. (4) Development of methods for correlation of image information from two-view DFMs. (5) Comparison of the detection and classification accuracy of the multiple-view fusion CAD system with the performance of the single-view CAD system by receiver operating characteristic (ROC) and free response ROC (FROC) analyses. (6) Adaptation of the computer vision techniques to the CAD system for DMs. (7) Adaptation of the multiple-view fusion methods to the CAD system for DMs.

We will develop novel regional registration methods for identifying corresponding lesions on craniocaudal (CC) and mediolateral oblique (MLO) views. The multiple image information will be fused with specially designed correspondence classifiers or fuzzy classification to reduce false positives and to improve lesion detection sensitivity. Multiple-view features of a lesion will be merged using neural networks or other classifiers for classification of malignant and benign lesions. In addition, new computer vision techniques will be developed in each of the four areas to improve the current methods. The techniques will be first developed for DFMs. The algorithms for DFMs will then be adapted to DMs, taking into account the differences in the imaging characteristics between DMs and DFMs. Databases of DFMs and DMs will be collected from our patient population with IRB approved protocol and extensive training and independent testing of the new CAD system will be performed. The test performance of the multiple-image correlation CAD algorithms for detection and characterization of lesions on DFMs will be compared with the one-

view approach on DFMs as well as the performances of CAD systems for DMs using ROC methodology.

DM or DFM not only has the potential to detect breast cancer in an early stage, it will also facilitate consultation via teleradiology in remote or rural regions where expert mammographers may not be readily available. An effective CAD system will be particularly useful for providing an additional on-site or remote second opinion. This will be highly relevant to women in the military, especially when they are stationed in remote areas. DM in combination with CAD will fully utilize the potential of mammography to improve the health care of women both in the military and in the general population.

(5) Body

This is the fourth year annual report of our project. In the project period (5/1/05-4/30/06), we have extended our investigations to both the CAD systems for DMs and DFMs, and performed a number of studies to develop the CAD system for breast cancer diagnosis. A summary of some of the important accomplishments follows.

(A) Collection of databases of digital mammograms and digitized film mammograms

We continue to collect the database of digital mammograms (DMs) with mammographic masses or clustered microcalcifications for the development of our computer-aided diagnosis (CAD) algorithms. We have collected about 280 cases containing more than 1120 mammograms. The patients were diagnosed with lesions in their mammograms during their normal clinical care, either by routine screening or by referral to our breast imaging clinic for evaluation. Most of the cases contained both DMs and screen-film mammograms.

As described in our previous reports, the digital mammograms are acquired with a GE Senographe 2000D full field digital mammography (FFDM) system. After acquisition, the digital image files are transmitted to the Siemens Archive which is the PACS system used in our department for storage of all clinical digital images. With Institutional Review Board (IRB) approval, we download the DMs from the Siemens Archive to our laboratory and digitize the film mammograms from the same patient. The film mammograms are digitized with a Lumiscan 85 laser scanner.

We have developed a database management program based on Microsoft Access to process the images downloaded to our system. For each mammogram file, all patient identifiers are first removed from the image header. The patient name is replaced with a code number. The image is then named by the code number. A record is generated in the database file for each image. The record keeps the code number, the lesion type, the view, and the exam date information for each case. If the pathology of the case is available, the malignant or benign information of the lesion is also entered. Each case in the database will be read by an experienced MQSA radiologist to mark the lesion location. For microcalcification cases, the radiologist measures the diameter of the cluster, and provides description of its distribution, morphology, and visibility of the microcalcifications. For mass cases, the radiologist measures the diameter of the mass, and provides description of its margin, shape, spiculated or non-spiculated, the visibility, and the density of the mass relative to that of the parenchyma. For all cases, the radiologist also provides BI-RADS description of the breast density and estimates the likelihood of malignancy of the lesion. These descriptions are entered into the database for each case as a reference for future analysis.

(B) CAD system for microcalcification detection on digital mammograms – comparison of detection accuracy on digitized film mammograms and digital mammograms

We are developing CAD systems to detect microcalcification clusters automatically on DMs and on DFMs. In this study, we compared the detection accuracy of the CAD systems using a data set of matched DMs and DFMs from the same patients.

Methods:

Our CAD system for microcalcification detection includes five stages: preprocessing, image enhancement, segmentation of microcalcification candidates, false positive (FP) reduction based on a convolution neural network (CNN), and regional clustering. The image processing and computer-vision techniques used in the CAD systems are the same for DMs and DFMs except that the preprocessing stage is different. For the DM CAD system, raw images are used as input to reduce the dependence of the system on specific manufacturer's proprietary preprocessing methods. An inverted logarithmic transformation is applied to the raw pixel values to convert the image pixel depth to 12-bit, similar to that of the DFMs. For both the DM and DFM CAD systems, the image is then subjected to an automated breast boundary segmentation algorithm. Further steps are only applied to the segmented breast area to reduce computation time. At the enhancement stage, the image is processed using a difference-image technique to enhance the signal-to-noise ratio (SNR) of the microcalcifications. Then potential signals are segmented from the image background using global and locally adaptive segmentation techniques. Rule-based classification is applied to the signal size, contrast and SNR to identify suspected individual microcalcifications. A convolution neural network (CNN) is trained to further exclude FP individual microcalcifications. A regional clustering procedure is then used to identify clustered microcalcifications. Finally, a trained LDA classifier is used to reduce FP microcalcification clusters from previous stage. The parameters and the feature classifiers are trained separately for the FFDM and DFM CAD systems.

Two data sets, one for DM and one for DFMs were collected. Each data set contained 96 cases with 192 images. All cases had two mammographic views: the CC view and the MLO view or the lateral (LM or ML) view. Twenty-eight cases contained biopsy-proven malignant clusters and 68 cases were benign.

Results:

The detection performance of the CAD system is evaluated by free response receiver operating characteristic (FROC) analysis. FROC curves can be compared on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the cluster on each mammogram is considered an independent true cluster. For case-based FROC analysis, the same cluster imaged on the two-view mammograms is considered to be one true object and the detection of either or both on the two views is considered to be a true-positive (TP). The FROC curves for the DM and DFM CAD systems are compared in Fig. 1. For case-based performance evaluation, the FFDM CAD system achieved detection sensitivities of 70%, 80%, and 90% at an average FP rate of 0.07, 0.16, and 0.63 per image, compared with an average FP rate of 0.15, 0.38, and 2.02 per image for the DFM CAD system. The difference was statistically significant ($p<0.05$). When the FP rates were estimated using mammograms negative for microcalcifications, the corresponding FP rates were 0.04, 0.11, and 0.33 per image for the FFDM CAD system, and 0.08, 0.14, and 0.50 per image for the DFM CAD system.

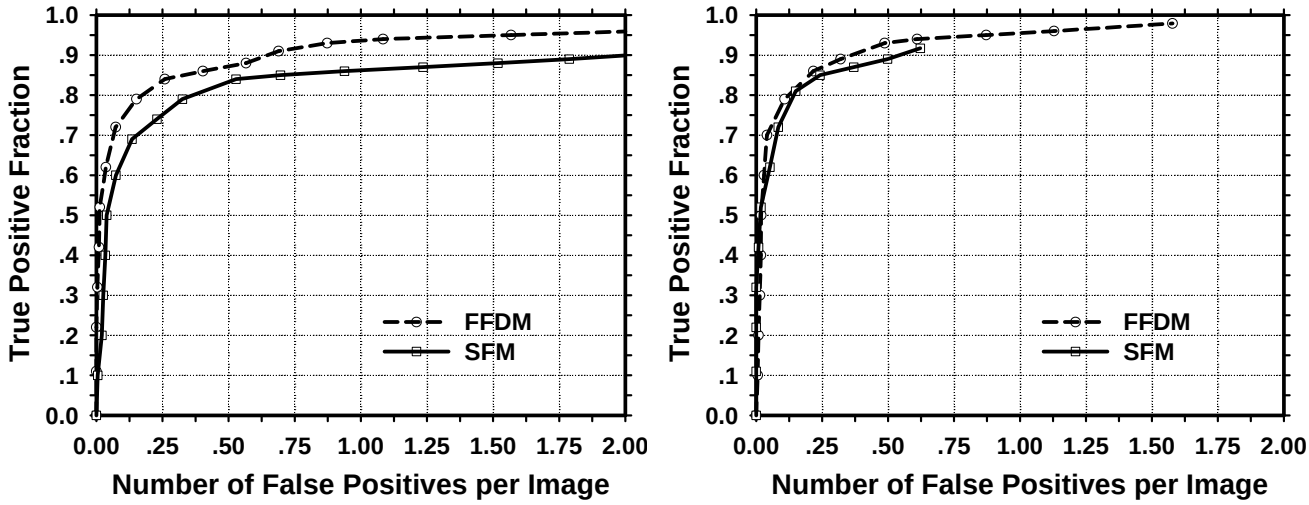


Fig. 1. Case-based FROC curves of the CAD systems for digital (FFDM) and screen-film (SFM) mammograms. The FP rates were estimated on mammograms with microcalcifications (left) and normal mammograms (right).

Conclusion:

The CAD system for microcalcification detection on DMs has a higher performance than that on DFMs in this data set. Since the sample size is small, it is unknown if the results can be generalized to unknown patient cases. Further study is underway to collect a larger data set and to improve the performance of the systems.

(C) CAD system for mass detection on mammograms – Two view information fusion

In screening mammography, radiologists utilize information from the CC view and the MLO view to confirm true mass and eliminate FPs. We are developing two-view fusion technique to combine the information from two mammographic views, thereby emulating radiologists' strategy in differentiating true masses and FPs.

Methods:

The fusion method used in this study is based on the assumption that the corresponding true mass on two different mammographic views will exhibit similarities in their geometric, morphological and textural features which are relatively invariant with respect to the imaging views. On the other hand, FPs detected by CAD system are expected to exhibit a lesser degree of similarity because they are usually objects formed by different normal tissues.

A schematic of our two-view CAD system is shown in Fig. 2. The single-view mass CAD system is first applied to each mammographic view independently. For a given detected object on one view, geometric pairing is performed using the nipple-to-object distance as the average radius of an annular region on the other view within which the detected objects can be paired with the given object. Similarity measures between each pair of objects are derived from the pairs of individual object features. The similarity features include morphological features, Hessian feature, correlation coefficients between the two paired objects and texture features. A similarity classifier is trained to distinguish between true and false pairs by merging the similarity features into a similarity score for

each object. The similarity score and the single-view object score of the object are then fused to form a final score for the object.

We randomly separated the cases in our data set into two independent equal sized data sets: 243 cases with 494 images and 232 cases with 478 images. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by FROC analysis. To evaluate the overall test performance, an average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from 2-fold cross validation.

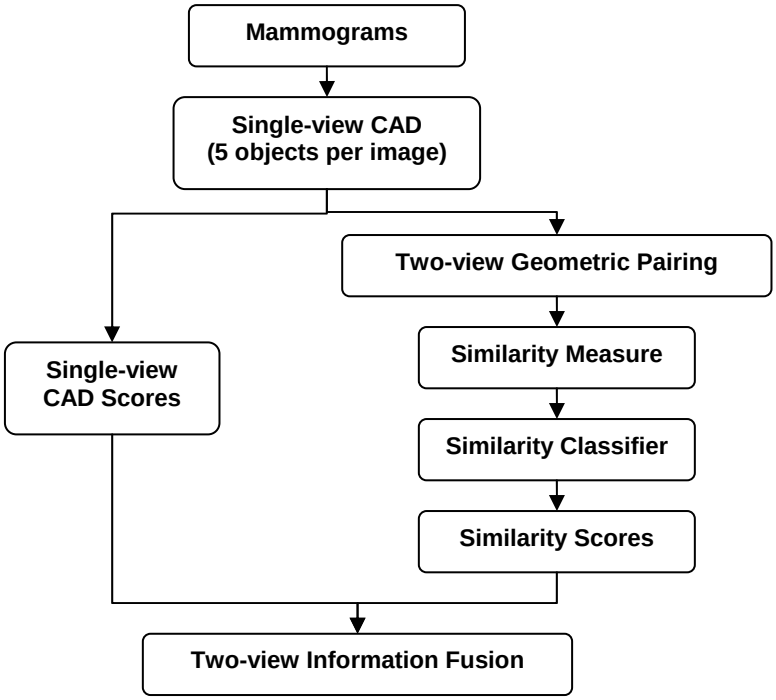


Fig. 2. Schematic of a two-view information fusion scheme for mass detection on mammograms.

Results:

When the single-view CAD system was applied to the test set, the FPs/image were 2.0, 1.5, and 1.2 at the case-based sensitivities of 90%, 85% and 80%, respectively. With the two-view CAD system, the FP rates were improved to 1.7, 1.3, and 1.0 FPs/image at the same case-based sensitivities. Fig. 3 shows the comparison of the test performance of the single-view CAD system and the two-view CAD systems by using image-based and case-based average FROC curves, respectively. The improvements in the test FROC curves for both subsets were statistically significant ($p<0.05$).

Conclusion:

Two-view information fusion is a promising approach to improving the performance for mass detection. Further work is underway to optimize the different stages of our two-view CAD system.

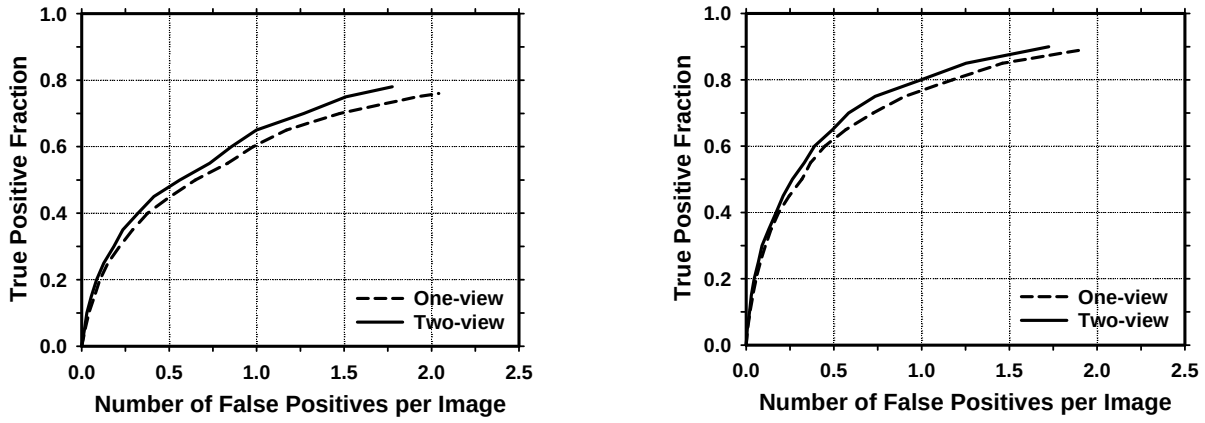


Fig. 3. Image-based (left) and case-based (right) average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. One-view: detection by the single-view CAD system. Two view: detection by using the two-view information fusion scheme.

(D) CAD system for mass detection on mammograms – Bilateral analysis for false positive reduction

Radiologists routinely compare density patterns on mammograms of the same view from the two breasts in mammographic interpretation. Asymmetric density can be caused by a new or developing lesion while symmetric density are more likely normal breast tissue. Bilateral comparison can therefore be used to detect new masses or to reduce FPs. We are developing computer-vision techniques to implement bilateral analysis in our CAD system for mass detection.

Methods:

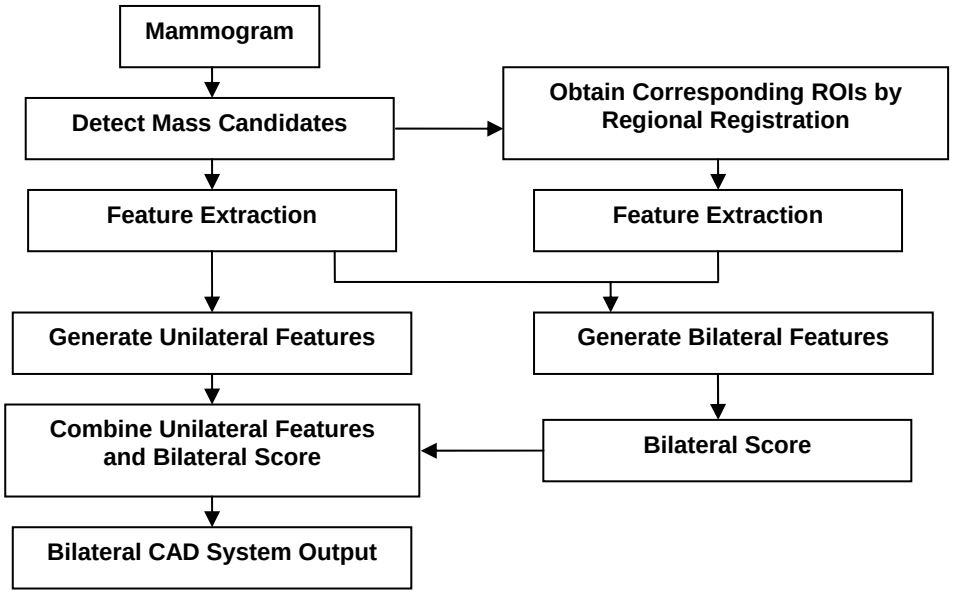


Fig. 4. Block diagram of the bilateral CAD system for FP reduction on mammograms.

A schematic of our bilateral analysis method is shown in Fig. 4. We first detect the mass candidates on each view by utilizing our unilateral CAD system. For each detected object, the regional registration technique is used to define a region of interest (ROI) that is "symmetrical" to the object location on the contralateral mammogram. Spatial gray level dependence matrices (SGLD) texture features and morphological features are extracted from both the ROI containing the detected object on a mammogram and its corresponding ROI on the contralateral mammogram. Bilateral features are then generated from the extracted unilateral features and a final bilateral score is formed as a new feature to differentiate symmetric from asymmetric ROIs. By incorporating the unilateral features of the mass candidates and their bilateral scores, a bilateral classifier was trained to reduce the FPs.

Results:

The FROC curves obtained from the unilateral and bilateral CAD systems are compared in Fig. 5. The bilateral CAD system achieved a case-based sensitivity of 70%, 80%, and 85% at 0.52, 0.83, and 1.05 FPs/image on the test data set. In comparison to the FP rates for the unilateral CAD system of 0.67, 1.11, and 1.69, respectively, at the corresponding sensitivities, the FP rates were reduced by 22%, 25%, and 37% with the bilateral symmetry information.

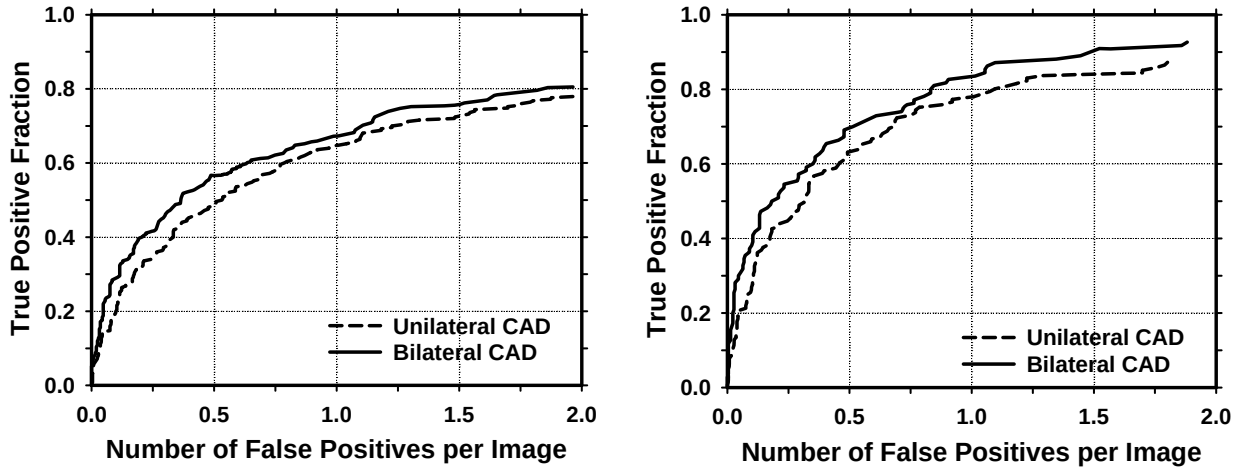


Fig. 5. (a) Image-based and (b) case-based FROC curves from the unilateral and the bilateral CAD systems.

Conclusion:

Our preliminary results demonstrate that the bilateral features can be utilized to differentiate the similarity and dissimilarity between tissues at corresponding locations in the bilateral views, and can be useful for improving the performance of a unilateral CAD system by further reducing the FPs. Further investigation is underway to improve the bilateral CAD system.

(E) Computer-aided mass detection on digital tomosynthesis mammograms (DTM) – Dependence on reconstruction image quality

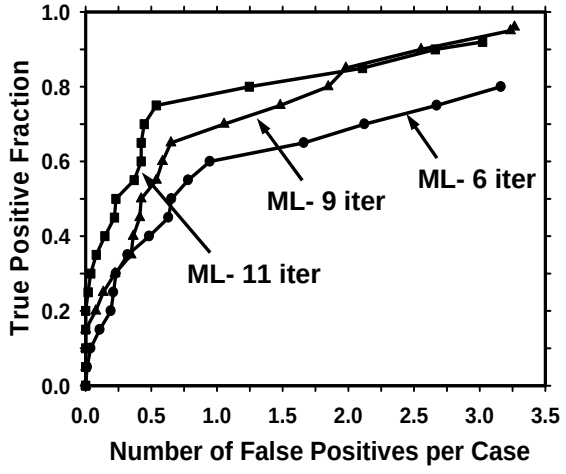
Digital tomosynthesis mammography (DTM) is a new modality that holds the promise of improving breast cancer detection. Although it is not included in our original proposal, we believe that this new modality in combination with CAD will be an exciting new direction for improving breast cancer detection and diagnosis. We thus performed a pilot study to investigate the feasibility of developing a CAD system for breast masses on DTMs. In the annual report last year, we described the image processing methods used in our CAD system. In this project period, we continue the feasibility study by evaluating the dependence of the performance of the CAD system on image quality of the reconstructed DTMs.

Methods:

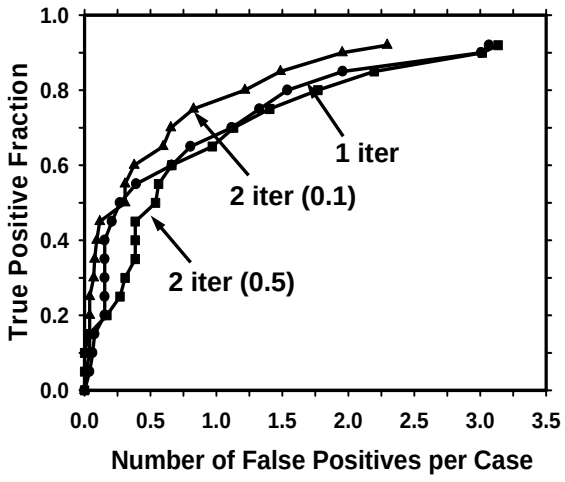
In our CAD system for DTMs, 3D gradient field analysis is first applied to the DTM volume for prescreening of mass candidates. Each mass candidate is segmented from the surrounding structured background by 3D region growing with adaptive thresholding of the radial gradient. Morphological, gray level, and texture features were then extracted from the segmented object, and a linear discriminant classifier with stepwise feature selection was designed to reduce FPs.

Our pilot data set consisted of 26 DTM cases including 23 masses (13 malignant) and 3 areas of architectural distortion (2 malignant). The cases were collected at the Massachusetts General Hospital (MGH) with IRB approval. The GE DTM prototype system at the MGH acquired 11 PVs of the compressed breast over a 50 deg arc in the MLO view. DTM slices were reconstructed at 1-mm slice spacing using an iterative maximum-likelihood (ML)-convex technique and a simultaneous algebraic reconstruction technique (SART). The image quality of the DTMs reconstructed using both methods depended on the number of iterations performed. We trained the CAD system using a leave-one-out resampling scheme. The system was optimized separately for DTM mammograms reconstructed at different iterations. The performances of the CAD systems at the different image construction conditions were compared using FROC analysis.

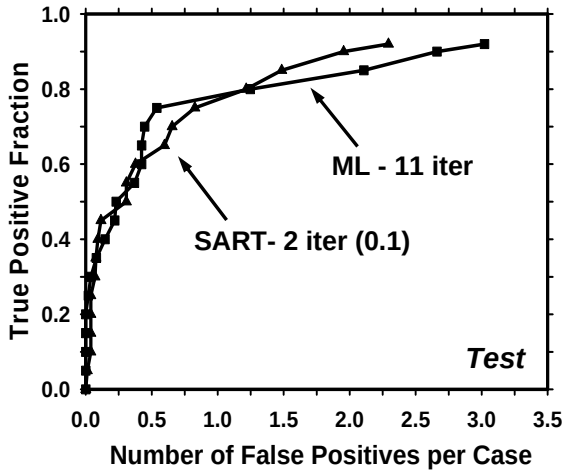
Results:



(a)



(b)



(c)

Fig. 6. FROC curves of the CAD system for DTMs reconstructed with different techniques: (a) Maximum likelihood-convex technique at 6 to 11 iterations, (b) simultaneous algebraic reconstruction technique at 1 to 2 iterations, and (c) comparison of the two techniques.

The FROC curves at selected conditions are shown in Fig. 6. For the ML technique, the FROC curve improved as the number of iterations increased from 6 to 11. For the SART, the best FROC curve was obtained with two iterations using a step size of 0.1 at the second iteration. When the highest FROC curves from the two reconstruction techniques are compared, both can achieve 80% sensitivity at about 1.2 FPs/case.

Conclusion:

CAD performance varied with the quality of the reconstructed DTM mammograms. It is important to evaluate the impact of reconstruction algorithms and their parameters on lesion detection accuracy for both CAD and human readers.

(6) Key Research Accomplishments

- Continue collection of a database of digital mammograms and digitized film mammograms for development of the CAD algorithms for both digital mammography and film mammography ------ (Task 1)

- Improve microcalcification detection CAD systems for digital mammograms and digitized film mammograms and compare the performance or the two CAD systems by FROC analysis ------- (Task 3(a), Task 6(a))

- Develop two-view information fusion to improve the accuracy of the CAD system for mass detection and evaluate the system performance by FROC analysis ------- (Task 2(a), Task 4(a), Task 6(a))

- Develop computer-vision techniques for bilateral analysis of mammograms to improve the CAD system for mass detection ------- (Task 5(a), Task 5(b))

- Explore computer-aided mass detection for digital tomosynthesis mammograms (DTM) and evaluate the effects of image reconstruction techniques on detection accuracy ------- (Task 2(a))

(7) Reportable Outcomes

As a result of the support by the PRMRP grant, we have conducted studies in CAD for mammography and published the results. The publications in this project year are listed in the following.

Peer-Reviewd Journal Articles:

1. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics 2005; 32: 2827-2838.

2. Chan HP, Wei J, Sahiner B, Rafferty EA, Wu T, Roubidoux MA, Moore RH, Kopans DB, Hadjiiski LM, Helvie MA. Computer-aided detection system for breast masses on digital tomosynthesis mammograms – Preliminary experience. Radiology 2005; 237:1075-1080.

Accepted for Publication in Peer-Reviewd Journals:

1. Hadjiiski LM, Sahiner B, Helvie MA, Chan HP, Roubidoux MA, Paramagul C, Blane C, Petrick N, Bailey J, Klein K, Foster M, Patterson S, Adler D, Nees A, Shen J. Computer-aided diagnosis of breast cancer in serial mammograms. Radiology.

2. Sahiner B, Chan HP, Roubidoux MA, Hadjiiski LM, Helvie MA, Paramagul C, Bailey JE, Nees A, Blane CE. Computer-aided diagnosis of malignant and benign breast masses in 3D ultrasound volumes: Effect on radiologists' characterization accuracy. Radiology.

3. Sahiner B, Chan HP, Hadjiiski LM, Helvie MA, Paramagul C, Ge J, Wei J, Zhou C. Joint two-view information for computerized detection of microcalcifications on mammograms. Medical Physics.

4. Ge J, Sahiner B, Hadjiiski LM, Chan HP, Wei J, Helvie MA, Zhou C. Computer aided detection of clusters of microcalcifications on full field digital mammograms. Medical Physics.

5. Zhang Y, Chan HP, Sahiner B, Wei J, Goodsitt MM, Hadjiiski LM, Ge J, Zhou C. A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis. Medical Physics.

Non-Peer-Reviewd Conference Proceeding Articles:

1. Zhou C, Hadjiiski LM, Paramagul C, Sahiner B, Chan HP, Wei J. Computerized pectoral muscle identification on MLO-view mammograms for CAD applications. Proc SPIE 5747; 2005: 852-857.

2. Hadjiiski LM, Chan HP, Sahiner B, Helvie MA, Roubidoux MA. Effects of the continuous and discrete confidence rating scales in ROC observer studies. Proc SPIE 5749; 2005: 1-7.

3. Ge J, Wei J, Hadjiiski LM, Sahiner B, Chan HP, Helvie MA, Zhou C, Ge Z. Computer aided detection of microcalcification clusters on full-field digital mammograms: multiscale pyramid enhancement and false positive reduction using an artificial neural network. Proc SPIE 5747; 2005: 806-812.

4. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Helvie MA, Roubidoux MA, Petrick N, Zhou C, Ge J. Computer aided detection of breast masses on mammograms: performance improvement using a dual system. Proc SPIE 5747; 2005: 9-15.

Conference Abstracts and Presentations:

1. Zhou C, Chan HP, Helvie MA, Wei J, Ge J, Hadjiiski LM, Sahiner B, Computerized mammographic breast density estimation on full field digital mammogram and digitized film mammogram. Presentation at the 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 27-December 2, 2005. RSNA Program 2005; 271.

2. Ge J, Chan HP, Hadjiiski LM, Wei J, Helvie MA, Zhou C, Sahiner B. Computer-aided detection system for clustered microcalcification: comparison of performance on full-field digital mammograms and digitized screen-film mammograms. Presentation at the 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 27-December 2, 2005. RSNA Program 2005; 701.

3. Chan HP, Wei J, Wu T, Sahiner B, Rafferty EA, Hadjiiski LM, Helvie MA, Roubidoux MA, Moore RH, Kopans DB. Computer-aided detection on digital breast tomosynthesis (DTM) mammograms: Dependence on image quality of reconstruction. Presentation at the 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 27-December 2, 2005. RSNA Program 2005; 269.

4. Hadjiiski LM, Chan HP, Sahiner B, Helvie MA, Roubidoux MA, Zhou C, Fully Automated Regional Registration and Classification of Corresponding Microcalcification Clusters on Serial Mammograms. Presentation at the 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 27-December 2, 2005. RSNA Program 2005; 270.

5. Ge J, Sahiner B, Chan HP, Hadjiiski LM, Helvie MA, Zhou C, Wei J, Zhang Y. Computer-Aided Detection of Clustered Microcalcifications on Full-Field Digital Mammograms: A Two-view Information Fusion Scheme for FP reduction. Poster presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

6. Wu YT, Hadjiiski LM, Wei J, Zhou C, Sahiner B, Chan HP. Computer-aided detection of breast masses on mammograms: bilateral analysis for false positive reduction. Presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

7. Sahiner B, Chan HP, Hadjiiski LM. Performance analysis of 3-class classifiers: Properties of the 3D ROC surface and the normalized volume under the surface. Presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

8. Hadjiiski LM, Drouillard D, Chan HP, Sahiner B, Helvie MA, Roubidoux MA, Zhou C. Characterization of Corresponding Microcalcification Clusters on Temporal Pairs of Mammograms for Interval Change Analysis - Comparison of Classifiers. Poster presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

9. Wei J, Sahiner B, Zhang Y, Chan HP, Hadjiiski LM, Zhou C, Ge J, Wu YT. Regularized Discriminant Analysis for Breast Mass Detection on Full Field Digital Mammograms. Poster presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

10. Zhang Y, Chan HP, Sahiner B, Wei J, Goodsitt MM, Hadjiiski LM, Ge J, Zhou C. Tomosynthesis Reconstruction with Simultaneous Algebraic Reconstruction Technique (SART) on Breast Phantom Data. Poster presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

11. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Helvie MA, Roubidoux MA, Zhou C, Ge J, Zhang Y. Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms. Presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February11-16, 2006.

12. Chan HP, Wei J, Sahiner B, Zhang Y, Hadjiiski LM, Helvie MA, Roubidoux MA, Zhou C, Ge J. Recent advances in computer-aided detection of breast cancer on mammograms. Poster presentation at the Peer Reviewed Medical Research Program (PRMRP) Investigators Meeting. Puerto Rico, May 1-4, 2006. Program book p.54.

(8) Conclusions

Under the support of this grant, we have investigated various computer-aided detection and diagnosis (CAD) methods for analysis of lesions on mammograms. We continue to collect a database of digitized film mammograms (DFMs) and a database of full field digital mammograms (DMs) that contain mammographic lesions from our breast imaging division in the Department of Radiology. The digital images include the manufacturer's processed images and unprocessed (raw) images. All collected cases are entered into our database management program that stores the coded case information to facilitate archiving and retrieval of the cases.

As discussed in the annual report last year, we continue to develop computer-vision techniques using DFMs in parallel with DMs. These techniques should be readily transferable between DFMs and DMs with minor modifications and retraining of the system parameters. In this project year, we compared the performance of microcalcification detection for the CAD system trained for DMs to that trained for DFMs. We found that both systems can provide high performances with the sensitivity slightly higher for DFMs.

We also continue to improve the CAD system for mass detection. We have developed two multiple-image analysis techniques for the mass detection system. The first is a two-view information fusion technique that combines the image information from the CC-view and the MLO-view

mammograms of the same breast. The second is a bilateral image analysis technique that compares the density patterns on left-breast and right-breast mammograms of the same view. The results of FROC analysis demonstrate that both techniques can increase the accuracy of the mass detection CAD system. These new image analysis techniques incorporate methods that are commonly used by radiologists in mammographic interpretation. The translation of human intelligence to computer vision is proven to be useful for improving the performance of the CAD system.

We continue to explore the development of a CAD system for digital tomosynthesis mammography (DTM). DTM is a new imaging modality that holds the promise to improve the detection and diagnosis of early breast cancer by reducing the camouflaging effect of dense breast tissue. In this project year, we evaluated the dependence of the performance of the CAD system on the image quality of reconstructed DTMs. We compared two reconstruction techniques – the iterative maximum likelihood and the simultaneous algebraic reconstruction techniques. It was found that both techniques can provide high-quality reconstructed DTMs if the parameters are properly chosen, and the detection accuracy of the CAD system for DTMs reconstructed using either techniques are comparable. However, the latter technique can reach high image quality with less number of iterations, thus reducing the computational costs.

In summary, we have investigated a number of areas in computer-aided detection of mammographic lesions. We have made progress in the six tasks proposed in the project in this and previous project years. This lays the strong foundation for us to continue the development and to improve the robustness of the CAD systems for digital mammograms and digitized film mammograms in the coming years.

(9) References

None.

(10) Appendix

Copies of the following publications are enclosed with this report.

Peer-Reviewd Journal Article:

1. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics 2005; 32: 2827-2838.

2. Chan HP, Wei J, Sahiner B, Rafferty EA, Wu T, Roubidoux MA, Moore RH, Kopans DB, Hadjiiski LM, Helvie MA. Computer-aided detection system for breast masses on digital tomosynthesis mammograms – Preliminary experience. Radiology 2005; 237:1075-1080.

Conference Proceedings:

1. Zhou C, Hadjiiski LM, Paramagul C, Sahiner B, Chan HP, Wei J. Computerized pectoral muscle identification on MLO-view mammograms for CAD applications. Proc SPIE 5747; 2005: 852-857.

2. Hadjiiski LM, Chan HP, Sahiner B, Helvie MA, Roubidoux MA. Effects of the continuous and discrete confidence rating scales in ROC observer studies. Proc SPIE 5749; 2005: 1-7.

3. Ge J, Wei J, Hadjiiski LM, Sahiner B, Chan HP, Helvie MA, Zhou C, Ge Z. Computer aided detection of microcalcification clusters on full-field digital mammograms: multiscale pyramid enhancement and false positive reduction using an artificial neural network. Proc SPIE 5747; 2005: 806-812.

4. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Helvie MA, Roubidoux MA, Petrick N, Zhou C, Ge J. Computer aided detection of breast masses on mammograms: performance improvement using a dual system. Proc SPIE 5747; 2005: 9-15.

Computer-aided detection of breast masses on full field digital mammograms

Jun Wei,[a] Berkman Sahiner, Lubomir M. Hadjiiski, Heang-Ping Chan, Nicholas Petrick, Mark A. Helvie, Marilyn A. Roubidoux, Jun Ge, and Chuan Zhou
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 7 February 2005; revised 16 June 2005; accepted for publication 16 June 2005; published 23 August 2005)

We are developing a computer-aided detection (CAD) system for breast masses on full field digital mammographic (FFDM) images. To develop a CAD system that is independent of the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM image as input and developed a multiresolution preprocessing scheme for image enhancement. A two-stage prescreening method that combines gradient field analysis with gray level information was developed to identify mass candidates on the processed images. The suspicious structure in each identified region was extracted by clustering-based region growing. Morphological and spatial gray-level dependence texture features were extracted for each suspicious object. Stepwise linear discriminant analysis (LDA) with simplex optimization was used to select the most useful features. Finally, rule-based and LDA classifiers were designed to differentiate masses from normal tissues. Two data sets were collected: a mass data set containing 110 cases of two-view mammograms with a total of 220 images, and a no-mass data set containing 90 cases of two-view mammograms with a total of 180 images. All cases were acquired with a GE Senographe 2000D FFDM system. The true locations of the masses were identified by an experienced radiologist. Free-response receiver operating characteristic analysis was used to evaluate the performance of the CAD system. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% at 0.72, 1.08, and 1.82 false positive (FP) marks/image on the mass data set. The FP rates on the no-mass data set were 0.85, 1.31, and 2.14 FP marks/image, respectively, at the corresponding sensitivities. This study demonstrated the usefulness of our CAD techniques for automated detection of masses on FFDM images. [DOI: 10.1118/1.1997327]

Key words: computer-aided detection, full field digital mammogram (FFDM), multiresolution image enhancement, gradient field analysis, stepwise linear discriminant analysis

I. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 and 55 years of age.[1] It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer.[2–4] Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially.[5–8] Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography.[9,10] Computer-aided lesion detection can be used during screening to reduce oversight of suspicious lesions that warrant further work-up. Computer-aided lesion characterization can assist in the estimation of the likelihood of malignancy of lesions by using image and/or other information during the diagnostic stage. The majority of studies to date show that CAD can improve radiologists' lesion detection sensitivity,[11–16] although Gur *et al.*[17] found that CAD had no significant effect on the radiologists in their academic setting when they averaged the results from both low-volume and high-volume radiologists. Further analysis of Gur's data by Feig *et al.*[18] indicated that the 17 low-volume radiologists in Gur's study achieved similar increase in sensitivity as reported in other studies. The outcome of CAD studies therefore depends on the study design and data analysis.

A number of investigators have reported CAD algorithms for detection of masses on mammograms. Their approaches to prescreening of mass candidates were based primarily on mass characteristics including: (1) asymmetric density between left and right mammograms,[19–22] (2) texture,[23,24] (3) spiculation,[25,26] (4) gray level contrast,[27–31] and (5) gradient.[32] Some of these approaches were refined with a combination of the mass characteristics. Feature classifiers were then used to further differentiate masses from normal breast tissues.

Most mammographic CAD algorithms developed so far are based on digitized screen-film mammograms (SFMs). In the last few years, full field digital mammographic (FFDM) technology has advanced rapidly because of the potential of digital imaging to improve breast cancer detection. Several manufacturers have obtained clearance from the FDA for clinical use. It is expected that FFDM detectors will provide higher signal-to-noise ratio (SNR) and detective quantum efficiency, wider dynamic range, and higher contrast sensitivity

than digitized mammograms. The spatial resolution of digital detectors may also be different from that of digitized SFMs even when their pixel pitches are equal. Li *et al.* investigated the performance of their CAD system on mass detection that was developed for SFMs and modified for FFDMs.[33] Their preliminary results on a small data set showed that it achieved 60% sensitivity at 2.47 false positives (FPs)/image. It is expected that proper adaptation based on the imaging characteristics of FFDMs and re-training of the CAD system with FFDMs would improve the performance. Because of the higher SNR and linear response of digital detectors, there is also a strong potential that more effective feature extraction techniques can be designed to optimally extract signals from the image and improve the accuracy of CAD. Several commercial CAD systems already obtained FDA approval for use with FFDMs. The commercial CAD systems generally reported similar performance on FFDMs and SFMs. However, their study was not reported in peer-reviewed journals so that the data set and algorithm are unknown. Recently, an assessment study[34] to compare the performance of two commercial and one research CAD system for SFMs showed that their mass detection sensitivities ranged from 67% to 72% and the FP rates ranged from 1.08 to 1.68 per four-view examinations. The differences in sensitivities were not significant whereas the differences in the FP rates were significant, depending on the examinations and CAD systems used.[34]

We have developed a CAD system for the detection of masses on SFMs in our previous studies.[30,35,36] We are developing a mass detection system for mammograms acquired directly by a FFDM system. In this study, we adapted our mass detection system developed for SFMs to FFDMs by optimizing each stage and retraining. In an effort to develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM as input and developed a multiresolution preprocessing scheme for image enhancement. A new technique was also designed for prescreening of mass candidates on the preprocessed images.

II. MATERIALS AND METHOD

A. Data sets

The mammograms were collected from patient files at the Department of Radiology with Institutional Review Board approval. Digital mammograms at the University of Michigan are acquired with a GE Senographe 2000D FFDM system. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of 100 μm $\times$100 μm and 14 bits per pixel. In this study, we used two data sets: a mass set containing FFDMs with malignant or benign masses and a no-mass set containing FFDMs without masses. The no-mass set was obtained from microcalcification cases collected for the development of our microcalcification CAD systems. The cases were included as normal, with respect to masses, only if they were verified to be free of masses by an experienced Mammography Quality Standards Act (MQSA) radiologist. Our mass detection system aims at application to screening mammography so that the mass cases, regardless of malignant or benign, are considered positive. All cases had two mammographic views, the craniocaudal view and the mediolateral oblique view or the lateral (LM or ML) view. The mass set contained 110 cases with a total of 220 images. The no-mass set contained 90 cases with a total of 180 images. The mass data set was used to estimate the detection sensitivity and the no-mass data set was used for estimating the FP rate. There were a total of 110 biopsy-proven masses in the mass data set. Eighty-seven of the masses were benign and 23 of the masses were malignant. A MQSA radiologist identified the locations of the masses, measured the mass sizes as the longest dimension seen on the two-view mammograms, provided descriptors of the mass shapes and mass margins, and also provided an estimate of the breast density in terms of BI-RADS category. Figure 1 shows the information of our data set which includes the distributions of mass sizes, mass shapes, mass margins, and breast density.

B. Methods

Our CAD system consists of five processing steps: (1) preprocessing by using multiscale enhancement, (2) prescreening of mass candidates, (3) identification of suspicious objects, (4) feature extraction and analysis, and (5) FP reduction by classification of normal tissue structures and masses. The block diagram for the detection scheme is shown in Fig. 2. These steps are described in more detail in the following.

We randomly separated the mass data set into two independent, equal sized subsets. Each subset contained 55 cases with 110 images. Cross validation was used for training and testing the algorithms. The training included selecting the preprocessing Laplacian pyramid reconstruction weights, adjusting the filter weights for prescreening and clustering, determining thresholds for rule-based classification, and selecting morphological and texture features and classifier weights. Once the training with one subset was completed, the parameters and all thresholds were fixed for testing with the other subset. The training and test subsets were switched and the training process was repeated. The overall detection performance was evaluated by combining the performances for the two test subsets. The trained algorithms with the fixed parameters were also applied to the no-mass mammograms to estimate the FP rate in screening mammograms.

1. Preprocessing

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to readers. The image preprocessing method used depends on the manufacturer of the FFDM system. To develop a CAD system that is less dependent on the FFDM manufacturer″s proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. We developed a multiscale preprocessing scheme for image enhancement.

Multiscale methods have been used for contrast enhancement of medical images. Since a multiscale method uses the information from a large number of frequency channels ex-

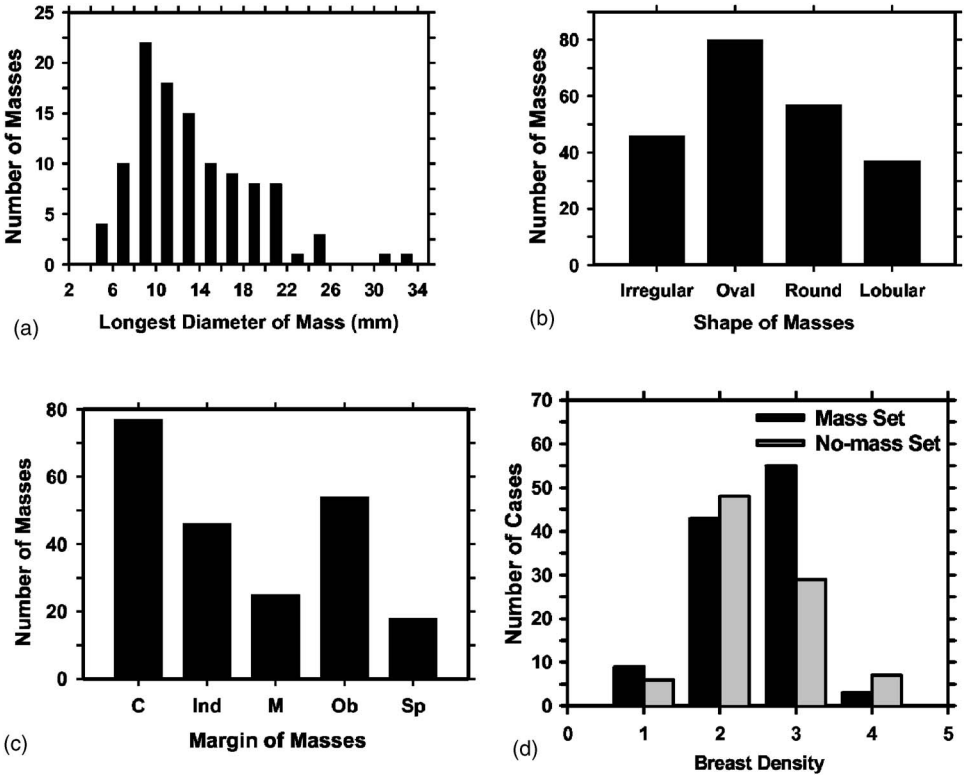


FIG. 1. The information of our mass data set: (a) distribution of mass sizes, (b) distribution of mass shapes, (c) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiqulated, (d) distribution of the breast density in terms of BI-RADS category estimated by a MQSA radiologist.

tracted from the image adaptively, it is more flexible and versatile than the commonly used enhancement methods, such as unsharp masking, which uses a small number of frequency channels. Two types of multiscale methods have been used as the preprocessing methods for the contrast enhancement of mammograms: the wavelet method and the Laplacian pyramid method.[37] A previous study has shown that, for the purpose of image enhancement, using a Laplacian pyramid method is advantageous compared to using the fast wavelet transformation which introduces visible artifacts.[38] In this project, therefore, we chose the Laplacian pyramid method as our preprocessing method.

A flowchart of our preprocessing method is shown in Fig. 3. In brief, the mammogram is first segmented automatically into the background and the breast region. Second, a logarithmic transform is applied to the breast image. The Laplacian pyramid method is used to decompose the breast image

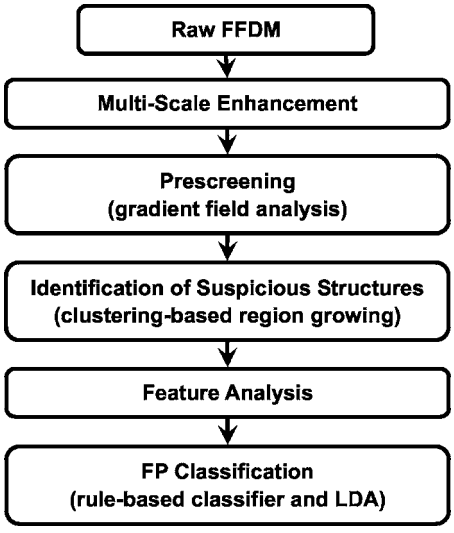


FIG. 2. Schematic diagram of our CAD system for mass detection on FFDM. The system is developed for screening mammography so that all masses, regardless of malignant or benign, are considered positive. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

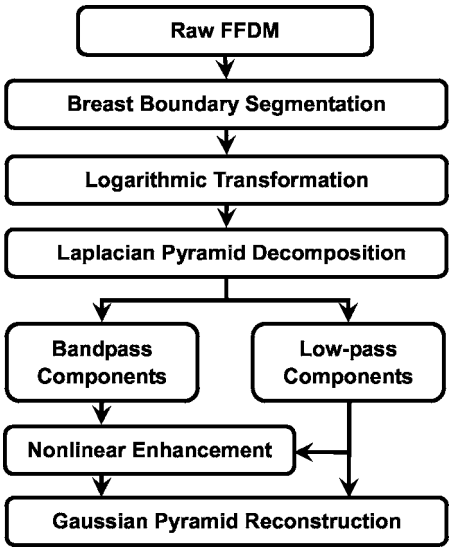


FIG. 3. Schematic diagram for the image preprocessing stage of our mass detection system, which includes breast boundary segmentation, logarithmic image transformation, and Laplacian pyramid multiscale enhancement.

into multiscales. A nonlinear weight function based on the pixel gray level from each of the low-pass components is designed to enhance the high-pass components.

Since the contrast between the breast and the background in a raw FFDM is high, a two-step algorithm was developed for the segmentation of breast region. First, Otsu's method[39] is used to calculate a threshold and binarize the original image. Second, an eight-connectivity labeling method is used to identify the connected regions below the threshold on the binary image. The region with the largest area will be considered to be the breast region.

Clinical mammograms are usually viewed in a negative mode of the raw images. In order to process an image with the same format as the clinical mammograms, we first use an inverted logarithmic function[40] to transform the raw data. A multiresolution method is then used to enhance the log-transformed image. The inverted logarithmic function for signal transfer can be expressed as

$$S_x = \ln\left(\frac{X_{\max}}{X}\right) \tag{1}$$

where X is the gray level of the raw data, $X_{\max}$ is the maximum of the 14 bit digital gray scale number (i.e., 16 383). The transformed image is then linearly scaled to 12 bit pixel values.

The Laplacian pyramid decomposition is a multiscale method that was first introduced as an image compression technique.[37] We previously evaluated the effect of Laplacian pyramid data compression on the detection of microcalcifications on digitized mammograms.[41] An illustration of a Laplacian decomposition tree is shown on the left-hand side of Fig. 4. The Laplacian pyramid is a sequence of error images $L_0, L_1, \ldots, L_n$. Each is the difference between two consecutive levels of the Gaussian pyramid $G_0, G_1, \ldots, G_n$, where G_0 is the original image. Each subsequent level of the Gaussian pyramid in the decomposition tree is generated by convolution of the image at the previous level with a 5×5 kernel, $w(m,n)$, that has weights of 0.4 at the center, 0.25 at the eight nearest neighbors of the center, and 0.05 at the 16 peripheral pixels, and then downsampled by a factor of 2, as described in Eq. (4). The decomposition of the image from level k to level $k+1$ can be expressed mathematically by

$$L_k = G_k - \text{Expand}(G_{k+1}), \tag{2}$$

where

$$\text{Expand}(G_{k+1}) = 4\sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)\cdot G_{k+1}\left(\frac{i-m}{2},\frac{j-n}{2}\right), \tag{3}$$

$$G_k(i,j) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)G_{k-1}(2i+m,2j+n). \tag{4}$$

The original image can be recovered by following the Gaussian reconstruction tree shown on the right-hand side of Fig. 4 if no enhancement is applied to the Laplacian pyramid. At a given level of the Gaussian reconstruction tree, the image is expanded (convolved and upsampled), as shown in Eq. (3), and then added to the Laplacian error image of the corresponding level. Details of the decomposition and reconstruction processes can be found in the literature.[37]

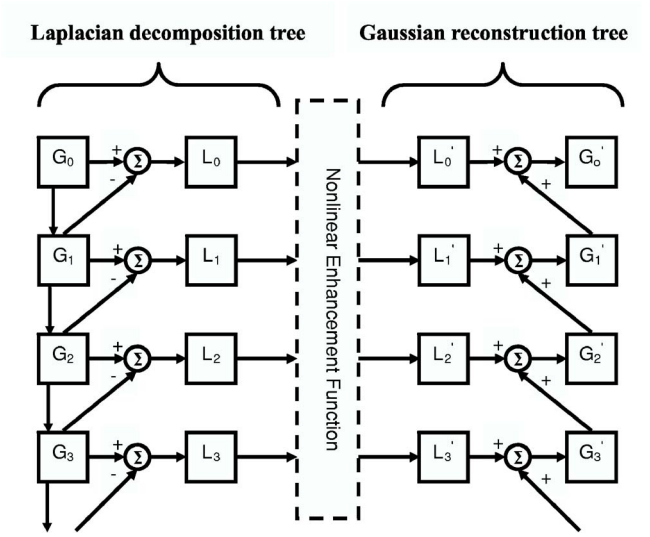


FIG. 4. Multiscale enhancement using the Laplacian pyramid decomposition method: Laplacian decomposition tree on the left-hand side and the Gaussian reconstruction tree on the right-hand side. The different levels of the Gaussian pyramid images are denoted by G_i, $(i=0,\ldots,n)$. The error images at different levels of the Laplacian pyramid are denoted by L_i, $(i=0,\ldots,n)$. The primed quantities G_i' and L_i' denoted the images at different levels after enhancement. $\sum$ denotes the summation operation. The image is downsampled by a factor of 2 when it goes down every level of the decomposition tree, and upsampled by a factor 2 when it moves up every level of the reconstruction tree.

We enhance the reconstructed image to facilitate mass detection. The image at each level of the Laplacian pyramid that corresponds to a bandpass image is mapped by a nonlinear function. In this study, we use a nonlinear function that incorporates the information from each bandpass image. A Gaussian pyramid expansion is then used to reconstruct the image from the low pass components and the enhanced bandpass components, as shown in Fig. 4. The reconstruction scheme is defined by

$$r(k) = \alpha\cdot\text{Expand}(G_{k+1}) + \beta\cdot(\text{Expand}(G_{k+1}))^p\cdot L_k, \tag{5}$$

where α, β, and p are constant values in the range of 0.2–2.0 experimentally chosen for each frequency level.

Figures 5(a) and 5(b) show an example of a GE raw image and its processed image provided by the GE FFDM system. The histograms of the raw image and the processed image are shown next to the corresponding images. An example of the processed image using our multiresolution enhancement method and the corresponding histogram are shown in Fig. 5(c).

2. Prescreening and segmentation of suspicious objects

In our previous CAD system developed for digitized SFMs, an adaptive density-weighted contrast enhancement (DWCE) filter[35] was developed for prescreening. Although

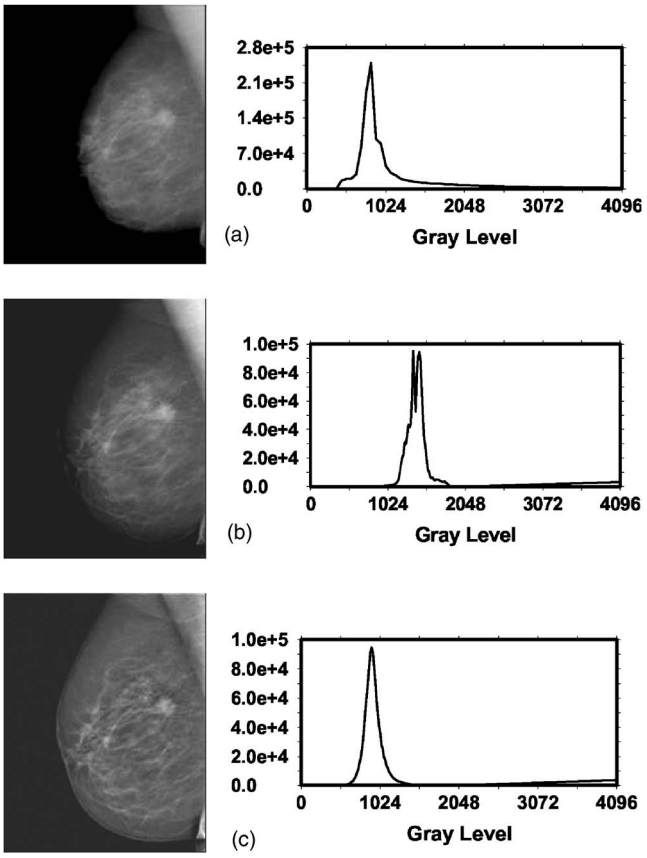


FIG. 5. An example of (a) GE raw image, (b) GE processed image, and (c) our processed image by using the Laplacian pyramid multiscale method. The gray level histogram of each image is also shown. The GE raw image has 14 bit gray levels but the histogram only plotted the lower 12 bits because very few pixels had gray levels higher than 4095.

the DWCE filter using the gray level information can identify the suspicious locations of masses on mammograms with high sensitivity, the prescreening objects often include a large number of enhanced normal breast structures.

In this study, we investigated the use of a new method that combines gradient field information and gray level information to detect mass candidates on FFDMs. Gradient field information is commonly used in computer vision or other fields to extract objects or intensity field distributions. Kobatake *et al.*[42] designed a filter, referred to as an iris filter, to calculate the convergence of gradient index around each pixel on SFMs which provided shape information for detection of masses. An extension of the iris filter, referred to as an adaptive ring filter, was developed by Wei *et al.*[43] for detection of lung nodules on chest x-ray images. In this study, we have developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis.

To reduce noise in the gradient calculation, the image is smoothed with a 4×4 box filter and subsampled to 400 μm $\times$ 400 μm. The gradient field analysis is applied to the smoothed image. At each pixel $c(i)$ within the breast, concentric annular regions centered at $c(i)$ with an average radius, $R(k)$, of k pixels from $c(i)$ and a radial width of 4 pixels are defined within a circular region of about 12 mm in radius. The gradient vector at each pixel $p(j)$ within an annular region is computed and the gradient direction is obtained by projecting the gradient vector to the radial direction vector from $c(i)$ to $p(j)$. The average gradient direction over an annular region at the average radius $R(k)$ is calculated as the mean of the gradient directions over pixels on three adjacent annular regions $R(k-1)$, $R(k)$, and $R(k+1)$. Finally, the gradient field convergence at $c(i)$ was determined as the maximum of the average gradient directions among all annular regions. A region of interest (ROI) of 256×256 pixels in the 100 μm $\times$ 100 μm images is identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method[44] in which the location of high gradient convergence is used as the starting point. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. Objects whose new gradient convergence is lower than 80% of the original value are rejected.

After prescreening, the suspicious objects are identified by using a two-stage segmentation method. First, the background-corrected ROI was weighted by a Gaussian function with $\sigma=256$ pixels. Then, a k-means clustering using the pixel values in a background-corrected image and a Sobel filtered image as features is used to find the object. Figures 6(a) and 6(b) show the initial detection locations and the grown objects, respectively, obtained by prescreening the mammogram shown in Fig. 5(c).

3. Feature extraction and FP reduction

FP classification in our mass detection system is accomplished by a three-stage classification scheme.[36,44] For each suspicious object, eleven morphological features are extracted. Rule-based classification and a linear discriminant analysis (LDA) classifier using all 11 morphological features as input predictor variables are trained to remove the detected structures that are substantially different from breast masses. The training data set alone was used for training the classification rules and the weights of the LDA classifier. After morphological classification, global and local multiresolution texture analyses[45] are performed in each remaining ROI by using the spatial gray level dependence (SGLD) matrix. Briefly, the wavelet transform is employed to decompose an ROI into three levels for global texture analysis. Thirteen types of texture features[44,46] are extracted from each ROI. Each feature is calculated at 14 pixel distances and 2 angular directions. A total of 364 features (13 texture measures $\times$ 14 distances $\times$ 2 directions) is extracted from global texture analysis. Local texture features are extracted from the local region containing the detected object (object region) and the peripheral regions within each ROI. A total of 208 features (104 features from the object region and 104

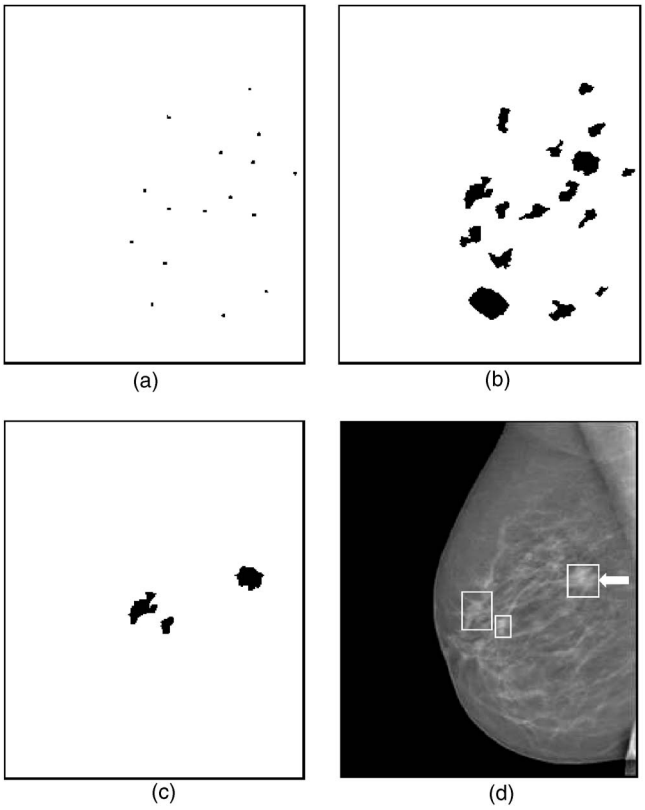


FIG. 6. An example demonstrating the processing steps with our CAD system: (a) object locations identified in prescreening, (b) identified suspicious objects, (c) detected objects after FP reduction, and (d) image superimposed with ROIs identifying the detected objects. The true mass is indicated by an arrow.

features from the peripheral regions) are extracted. The third-stage FP reduction using the texture features is described next.

4. Texture classification of masses and normal tissue

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection with stepwise LDA was applied. At each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. The optimization procedure used a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that had high correlation with the features already in the selected pool. Since the appropriate values of F_{in}, F_{out}, and T were unknown, we examined a range of F_{in}, F_{out}, and T values using an automated simplex optimization method. For a given combination of F_{in}, F_{out}, and T values, the algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To evaluate the classifier performance, the test discriminant scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology.[47] The discriminant scores of the mass and normal tissue were used as the decision variable in the LABROC program, which fits a binormal ROC curve based on maximum likelihood estimation. The accuracy for classification of mass and normal tissue was evaluated as the area under the ROC curve, A_z. The test A_z for the left-out cases in the leave-one-out resampling within the training subset was used as a figure of merit to guide the simplex algorithm to search for the best set of F_{in}, F_{out}, and T values within the parameter space. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased.[48] However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. In order to obtain a single trained classifier to apply to the test subset, a final stepwise feature selection was performed with the entire training subset and a set of F_{in}, F_{out}, and T thresholds chosen from the output of simplex training process. This set of F_{in}, F_{out}, and T thresholds was chosen based not only on the test A_z values, which were generated when the simplex procedure was searching through the parameter space, but also on the average number of features selected. The appropriate thresholds were chosen as a balance between keeping the number of selected features small and a relatively high classification accuracy by LDA. The chosen thresholds were then applied to the entire training subset to obtain the final set of features using stepwise feature selection and estimate the weights of the LDA. The LDA classifier with the selected feature set was then fixed and applied to the test subset. The test subset was independent of the training subset as described in Sec. II B 2 and was not used in the above-described leave-one-case-out classifier training process.

5. Evaluation methods

The detected individual objects were compared with the "truth" ROI marked by an experienced radiologist. A detected object was scored as true positive (TP) if the overlap between the bounding box of the detected object and the truth ROI was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study.[36] The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 220 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to 110 masses. Figure 6(c) shows an example of the final detected objects and Fig. 6(d) shows the locations of these objects superimposed on the mammogram.

To evaluate the effect of the preprocessing methods on mass detection, we also trained a CAD system using the GE

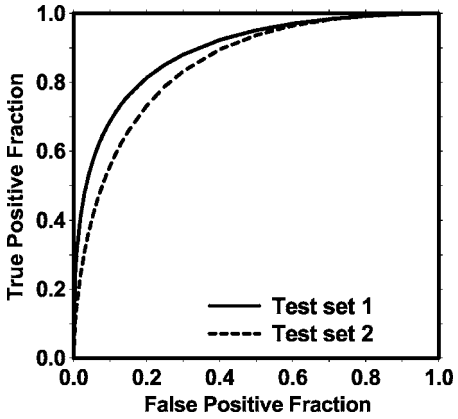


FIG. 7. The test ROC curves from the two independent mass subsets. The LDA classifiers using text features achieved an A_z value of 0.89±0.02 for test subset 1 and 0.85±0.02 for test subset 2 in the classification of mass and normal breast tissues.

processed image as input. This CAD system used the same methods as those described earlier for the raw images except that the Laplacian pyramid preprocessing step was not applied to the GE processed image, and that the prescreening and feature classifiers were retrained specifically for the GE processed images to obtain the best performance. The training and test subsets contained the same corresponding cases as for the raw image subsets. The training and testing were performed using the above-described cross validation method. The performance of the CAD system using the GE processed images was quantified by the average test FROC curve and compared with that using the raw images.

III. RESULTS

With raw images as input and Laplacian pyramid enhancement, our CAD system using the two-stage gradient field analysis detected 92.7% (204/220) of the masses with an average of 18.9 (4152/220) objects/image at the prescreening stage, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone. After FP reduction using the rule-based and linear classifier based on morphological features, there were a total of 3412 mass candidates (15.5 objects/image) at a sensitivity of 90.5% (199/220).

The texture-based LDA classifier for FP reduction was designed with stepwise feature selection and simplex optimization. The most effective subset of features from the available feature pool was selected for each of the training subsets during the training procedure. Twenty (11 global and 9 local) and 19 (12 global and 7 local) texture features were selected from the two independent training subsets, respectively. The test ROC curves are shown in Fig. 7. The training A_z values of the LDA classifier on the two training subsets were 0.87±0.02 and 0.88±0.01, respectively. The classifiers achieved A_z values of 0.89±0.02 and 0.85±0.02 on the independent test subsets, respectively. Figure 8 shows the FROC curves for the two test subsets after FP reduction with the corresponding trained LDA classifiers. An average FROC curve was derived from these two FROC curves by averaging the FP/images at the corresponding sensitivities. This average test FROC curve is plotted in Fig. 9 for comparison with the other FROC curves, described next.

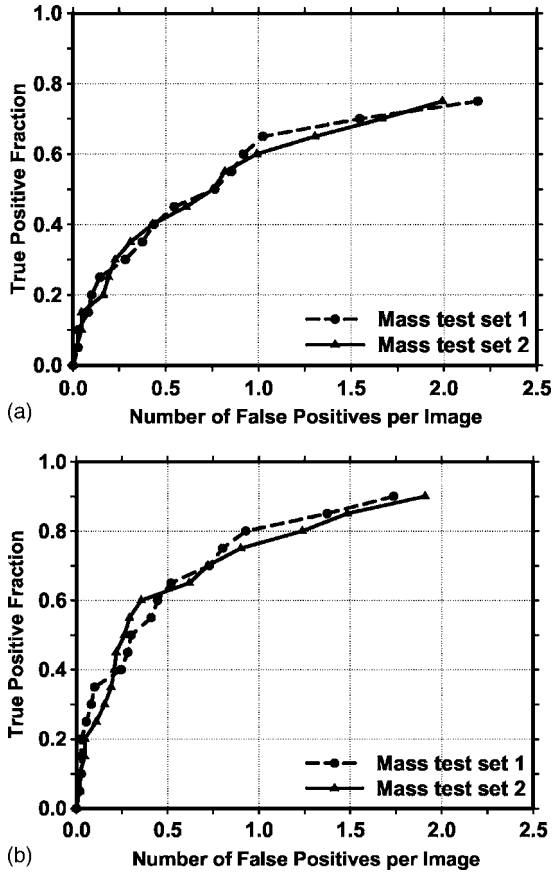


FIG. 8. The test FROC curves from the two independent mass subsets for the CAD system using the raw images as input and processed with the Laplacian pyramid method. The FP rate was estimated from the mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

In addition to using the mass data set containing 110 cases for the cross validation training and testing, we used a no-mass data set containing 90 cases with 180 images to evaluate the FP detection rate in normal cases. Since two sets of trained parameters were acquired as a result of the cross validation training, we applied the two trained CAD systems separately to the no-mass data set for FP detection. The number of FP marks produced by the algorithm was determined by counting the detected objects on these normal cases only. The mass detection sensitivity was determined by counting only the abnormal objects on each of the test mass subsets. The combination of the sensitivity from each of the test mass subsets and the FP rate from the normal data set at the corresponding detection thresholds resulted in a test FROC curve. The two test FROC curves were then averaged, as described earlier, to obtain an overall FROC curve quantifying the test performance of the CAD system. Figures 9(a) and 9(b) show the comparison of the average FROC curves with the FP rates estimated from the two data sets. The test FROC curve with the FP rate estimated from the no-mass data set showed a case-based detection sensitivity of 70%,

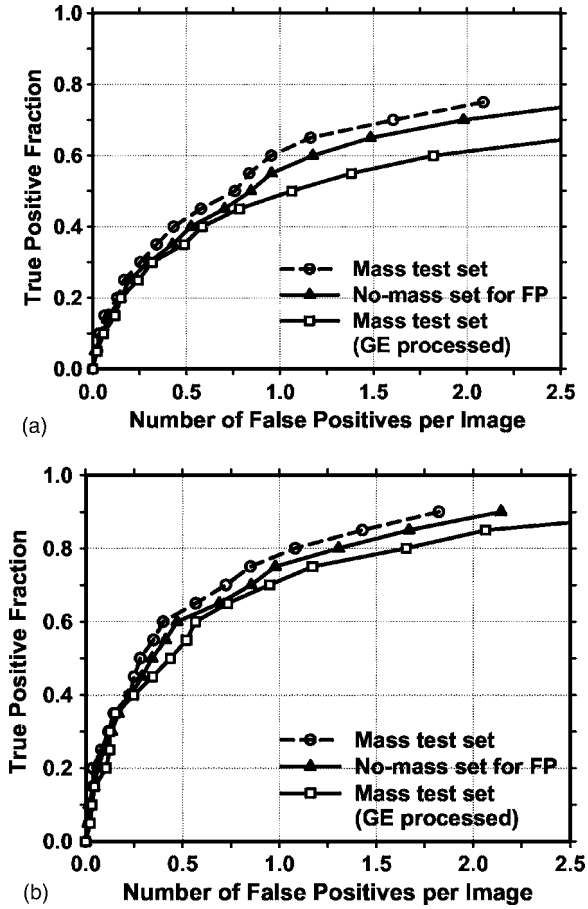


FIG. 9. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input, with the FP rate estimated from the mammograms with masses, (2) the CAD system using raw images as input, with the FP rate estimated from the normal mammograms without masses, and (3) the CAD system using GE processed images as input, with the FP rate estimated from the GE processed mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

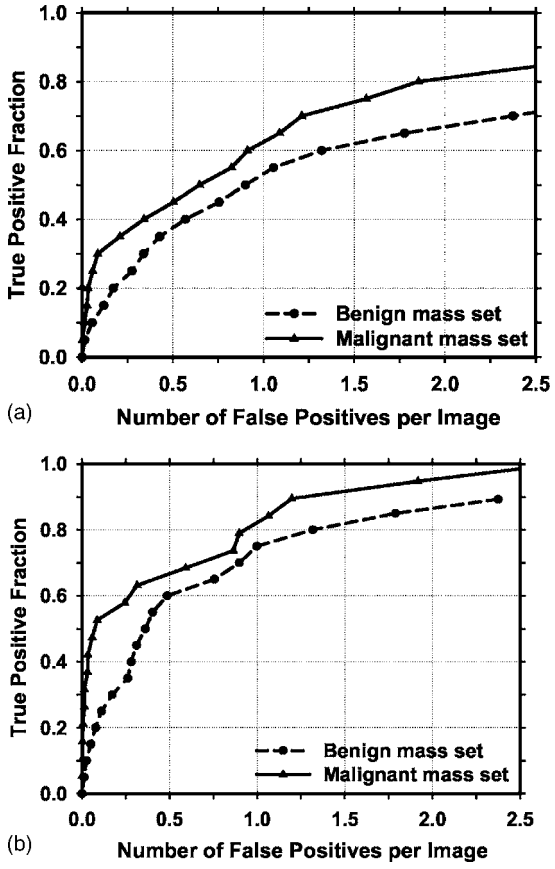


FIG. 10. Comparison of the average test FROC curves for the malignant and benign mass sets. The CAD system using raw images as input was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

80%, and 90% at 0.85, 1.31, and 2.14 FP marks/image, which are slightly higher than the FP rates of 0.7, 1.1, and 1.8 marks/image, respectively, estimated from the mass data set. Since our mass detection algorithm limits the maximum number of output marks to be 3 at the final stage, the FP marker rates will be slightly higher if the detection is performed in no-mass images. However, many images do not reach the maximum of 3 marks so that the difference in the FP marker rate between the mass and no-mass set is less than one. We also analyzed the detection accuracy of the system for malignant and benign masses separately. Figures 10(a) and 10(b) show the average FROC curves for detection of malignant and benign masses.

The average test FROC curves of the CAD system using the GE processed images as input were compared to those of the CAD system using raw images as input and Laplacian pyramid multiscale preprocessing as shown in Fig. 9. The FROC curves were plotted as the detection sensitivity as a function of the number of FP marks per image on the mass data set. The CAD system using the GE processed images as input achieved a case-based sensitivity of 70%, 80%, and 90% at 0.9, 1.6, and 3.1 FP marks/image, respectively, compared with 0.7, 1.1, and 1.8 FP marks/image on the CAD system using raw images as input.

IV. DISCUSSION

Several FFDM systems have been approved for clinical applications. It is important to develop a CAD system that can easily be adapted to images acquired by FFDM systems from different manufacturers. In this study, we are developing a CAD system that uses the raw FFDMs as the input. Since digital detectors generally have a linear response to x-ray exposure, the raw pixel values are a linear function of the absorbed x-ray energy in the detector. The signal range between different digital detectors can therefore be normalized linearly with respect to each other. Although the spatial resolution and noise properties of the images from different detectors are still different, the use of raw images already reduces one of the major differences between mammograms from different FFDM systems. For preprocessing of the raw images, we developed a multiresolution enhancement method. An example of a typical mammogram processed by the GE method and our method is compared in Fig. 5. As seen from this example, the enhancement of mammographic structures was stronger for our processed image than for the

TABLE I. Estimation of the statistical significance in the difference between the FROC performance of the CAD system using the FFDM raw images as input and processed with our Laplacian pyramid method and that of the CAD system using GE processed images as input. The FROC curves with the FP rates obtained from the no-mass data set (Fig. 9) were compared.

	A_1 (AFROC)			FOM (JAFROC)		
	Test subset 1	Test subset 2	p values	Test subset 1	Test subset 2	p values
Raw+LP processed	0.44	0.39	0.012	0.46	0.41	0.006
GE processed	0.37	0.31	0.0009	0.39	0.34	0.012

GE processed image. From a comparison of their histograms, it was found that the two histograms are very similar except for the average gray level.

For the evaluation of the effect of the preprocessing methods on computerized mass detection, we observed that our Laplacian pyramid preprocessing method provided higher detection accuracy than the GE processing method. As shown in Fig. 5, the Laplacian pyramid preprocessing method applies a stronger edge enhancement to the image than the GE method. Our preprocessing method aims at enhancing the image structures for computer vision whereas the GE processing method was designed to enhance the image for human visual interpretation. The stronger enhancement used for preprocessing the raw images appeared to improve the accuracy of the computer in detecting the masses.

Currently, there is no established statistical analysis method for testing the significance of the difference between two FROC curves generated by a CAD system. Chakraborty *et al.* proposed using an alternative free-response ROC (AFROC) method[49] to transform the FROC data to AFROC data, to which the curve fitting software and statistical significance tests for ROC analysis can then be applied and demonstrated its application to human observer performance rating data. In the AFROC method, false-positive images (FPIs) instead of FPs per image are counted. The confidence rating of a FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions are made on the same image. We applied the AFROC method to evaluate the differences in pairs of our FROC curves that used the no-mass set for estimation of the FP rates. The ROCKIT software developed by Metz *et al.*[47] was used to analyze the AFROC data. The comparison of A_1 and p values is summarized in Table I. The area under the fitted AFROC curve (A_1) was 0.44 and 0.39, respectively, on mass test subsets 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.37 and 0.31, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the fitted AFROC curve for our processed images and that for the GE processed images was statistically significant ($p<0.05$) for both test subsets. However, all four fitted AFROC curves deviated systematically from the AFROC data (see two examples plotted in Fig. 11 for the test subset 1). It is uncertain whether the AFROC method is applicable to our FROC data and thus whether the statistical significance testing is valid.

More recently, Chakraborty *et al.*[50] proposed a JAFROC method and provided software to estimate the statistical significance of the difference between two FROC curves. We also applied the JAFROC analysis to the two pairs of FROC curves. The figure-of-merit (FOM) from the output of the JAFROC software was 0.46 and 0.41, respectively, on mass test subsets 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.39 and 0.34, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the FOM for our processed images and that for the GE processed images was again statistically significant ($p<0.05$). The FOM values were about 0.02 higher than the corresponding A_1 values. The JAFROC software did not provide a fitted curve or a goodness-of-fit indicator in the output so that it is not known whether this model fits our FROC data better than the AFRPC method. Although both methods indicate that the improvement in the FROC performance using our Laplacian pyramid processed images is statistically

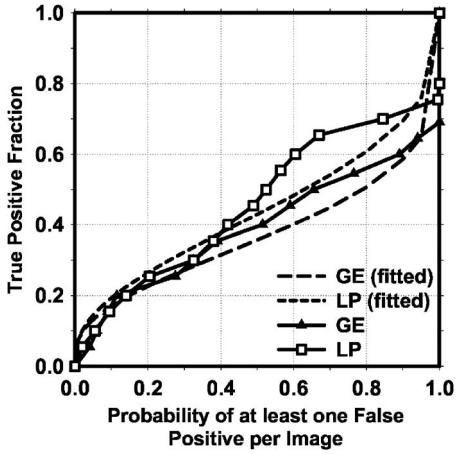


FIG. 11. Comparison of alternative free-response receiver operating characteristic (AFROC) curves. The raw curves were transformed from the FROC curves of mass detection on test subset 1 using either the raw images as input and processed with the Laplacian pyramid method (LP) or the GE processed images as input. The FP rate was estimated from the mammograms without masses. The fitted AFROC curves were obtained by applying the ROCKIT program to the transformed AFROC data.

significant, further investigations are needed to study whether these models are valid for analyzing the FROC performance of CAD systems.

The prescreening technique is an important task in a CAD system. A number of researchers have developed methods for detection of suspicious masses on SFMs and CRs. The previous methods produced between 10 to 30 FPs/image for a mass detection sensitivity of approximately 90%. However, it is difficult to compare the effectiveness of the different methods because of the differences in the image recording systems and in the data sets. In this study, we developed a new method that combines gradient field information, which was originally developed for the detection of lung nodules on chest x-ray images,[43] and gray level information[44] for prescreening mass candidates on the FFDMs. The new method produced 18.9 objects/image at 93% sensitivity in the prescreening step, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone.

The texture features in this study were extracted by using the SGLD matrix. A total of 572 features were included in our initial feature pool. These features were also used by our CAD system previously developed for SFMs. An average number of 19.5 features were selected by using a stepwise feature selection method. The A_z values for the LDA classifiers were 0.87 ± 0.02 and 0.88 ± 0.01 on the two training subsets, and 0.89 ± 0.02 and 0.85 ± 0.02 on the test subsets, respectively. The slightly higher test A_z from the first test subset than the A_z from its training subset may indicate that some relatively easy cases were assigned, by chance, to that test set during random partitioning. We also investigated if other features could improve the performance of our CAD system. The different feature spaces that we examined included features extracted from principal component analysis applied to the ROI image, run length statistics texture features extracted from the ROI images, and combination of one or both of these feature spaces with the SGLD feature space. However, the test results showed that a LDA classifier designed in the SGLD feature space alone provided the best performance. Although this was found to be true for both our CAD mass detection system for SFMs developed previously and the current system for FFDMs, it is still difficult to conclude that the SGLD features are the best feature set for classification between breast masses and normal tissues. One major concern of the SGLD feature space is that the dependence of the feature values on the pixel pair distance and angular direction leads to a feature pool with a large number of features. Some features in such a large feature space may provide good performance in classification of masses and normal structures by chance. We attempted to alleviate this problem by using an independent test set to evaluate the classifier performance. However, since we chose the overall system parameters with the knowledge of the performance for the test sets, the evaluation would still amount to validation rather than true testing. We have verified that our CAD system for SFMs can achieve reasonable performance in a true independent data set[36] and a prospective pilot clinical trial.[16] The performance of the current CAD system for FFDMs will have to be evaluated similarly when independent data sets become available.

The detection performance of a CAD system for malignant masses is more important than its performance for all masses. Figures 10(a) and 10(b) indicate that the sensitivity of the system is higher for malignant masses than for benign masses. This is consistent with our observation in previous studies of our CAD system for digitized SFMs.[36] However, since our current data set contained only 23 malignant cases, there will be large statistical uncertainty in the evaluation of sensitivity in this subset. A larger data set is being collected for comparing the detection performances of the CAD system between malignant and benign masses and also for the purpose of classifying malignant and benign masses. Furthermore, CAD algorithms developed for SFMs have been proven to be useful as a second opinion to assist radiologists in mammographic interpretation. Because of the higher SNR and linear response of digital detectors, there is also a potential that FFDMs can improve the sensitivity of breast cancer detection, especially in dense breasts. Several studies have been or are being conducted to compare FFDM with SFM in screening cohorts. It is also important to compare the performance of CAD systems between FFDMs and SFMs. A study is under way to compare the performance of the two systems on pairs of FFDM and SFM obtained from the same patients.[51]

V. CONCLUSION

Several FFDM systems have been approved for clinical applications. It is important to develop CAD systems for breast cancer detection in FFDM. In this work, we developed a CAD system that uses the raw FFDMs as the input. A multiresolution Laplacian pyramid enhancement method was devised to preprocess the raw FFDMs. A new prescreening method that combined gradient field analysis with gray level information was developed to identify mass candidates. Rule-based and LDA classifiers in a feature space which consisted of morphological features and SGLD texture features were designed to differentiate masses from normal tissues. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% with an estimate of 0.85, 1.31, and 2.14 FP marks/image, respectively, on normal cases. The results indicate that our mass detection CAD scheme can be useful for detecting masses on FFDMs. Studies are under way to further optimize the processing parameters, the feature extraction, and the classifiers for FP reduction. Comparison of mass detection performance of our CAD system for FFDMs and that for SFMs is also in progress.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA95153, U. S. Army Medical Research and Material Command Grant Nos. DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be

inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

[a]Electronic mail: jvwei@umich.edu

[1]American Cancer Society, www.cancer.org 2004, "Statistics for 2004."
[2]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," Cancer (N.Y.) **75**, 1619–1626 (1995).
[3]S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," Am. J. Roentgenol **171**, 29–33 (1998).
[4]B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," Cancer (N.Y.) **91**, 1699–1703 (2001).
[5]B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey, and C. J. Harris, "Mammogram interpretation by physician assistants," Am. J. Roentgenol. **149**, 907–911 (1987).
[6]C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists—Findings from a national sample," Arch. Intern Med. **156**, 209–213 (1996).
[7]R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," Radiology **219**, 192–202 (2001).
[8]J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," J. Natl. Cancer Inst. **95**, 1384–1393 (2003).
[9]F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," Orlando, FL.
[10]C. J. Vyborny, "Can computers help radiologists read mammograms?," Radiology **191**, 315–317 (1994).
[11]H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).
[12]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**, 554–562 (2000).
[13]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology **220**, 781–786 (2001).
[14]R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," Am. J. Roentgenology **181**, 687–693 (2003).
[15]S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," Radiology **232**, 578–584 (2004).
[16]M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey *et al.*, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection—A pilot clinical trial," Radiology **231**, 208–214 (2004).
[17]D. Gur, J. H. Sumkin, and H. E. Rockette, "Response to Re: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1261 (2004).
[18]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1260–1261 (2004).
[19]F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," Radiology **89**, 211–115 (1967).
[20]J. L. Semmlow, A. Shadagopappan, L. V. Ackerman, W. Hand, and F. S. Alcorn, "A fully automated system for screening mammograms," Comput. Biomed. Res. **13**, 350–362 (1980).
[21]T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," Comput. Biomed. Res. **24**, 273–295 (1991).
[22]F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," Med. Phys. **18**, 955–963 (1991).
[23]C. Kimme, B. J. O'Laughlin, and J. Sklansky, *Automatic Detection of Suspicious Abnormalities in Breast Radiographs*, Data Structures, Computer Graphics and Pattern Recognition (Academic, New York, 1975).
[24]W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," Radiology **191**, 331–337 (1994).
[25]S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," Comput. Biomed. Res. **25**, 218–237 (1992).
[26]N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," IEEE Trans. Med. Imaging **15**, 611–619 (1996).
[27]S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Imaging **8**, 377–386 (1989).
[28]D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," IEEE Trans. Med. Imaging **9**, 233–241 (1990).
[29]A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," IEEE Trans. Med. Imaging **13**, 725–740 (1994).
[30]N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," Med. Phys. **23**, 1685–1696 (1996).
[31]B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," Acad. Radiol. **2**, 959–966 (1995).
[32]H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computer detection of malignant tumors on digital mammograms," IEEE Trans. Med. Imaging **18**, 369–378 (1999).
[33]L. Li, R. A. Clark, and J. A. Thomas, "Computer-aided diagnosis of masses with full-field digital mammography," Acad. Radiol. **9**, 4–12 (2002).
[34]D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided detection performance in mammographic examination of masses: Assessment," Radiology **233**, 418–423 (2004).
[35]N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," IEEE Trans. Med. Imaging **15**, 59–67 (1996).
[36]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," Radiology **224**, 217–224 (2002).
[37]P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **COM-31**, 337–345 (1983).
[38]S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale constrast enhancement for radiographies: laplacian pyramid versus fast wavelet transform," IEEE Trans. Med. Imaging **21**, 343–353 (2002).
[39]N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979).
[40]A. Burgess, "On the noise variance of a digital mammography system," Med. Phys. **31**, 1987–1995 (2004).
[41]H. P. Chan, S. C. B. Lo, L. T. Niklason, D. M. Ikeda, and K. L. Lam, "Image compression in digital mammography: Effects on computerized detection of subtle microcalcifications," Med. Phys. **23**, 1325–1336 (1996).
[42]H. Kobatake and S. Hashimoto, "Convergence index filter for vector fields," IEEE Trans. Image Process. **8**, 1029–1038 (1999).
[43]J. Wei, Y. Hagihara, and H. Kobatake, "Detection of rounded opacities on chest radiographs using convergence index filter," ICIAP 99, Venice, 27–29 September, 1999, pp. 757–761.
[44]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. **26**, 1642–1654 (1999).

[45]D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," Med. Phys. **24**, 903–914 (1997).
[46]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973).
[47]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. **17**, 1033–1053 (1998).
[48]B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," Med. Phys. **27**, 1509–1522 (2000).
[49]D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," Med. Phys. **16**, 561–568 (1989).
[50]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[51]J. Wei, B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, and M. A. Helvie, "Computer aided diagnosis system for mass detection: Comparison of performance on full-field digital mammograms and digitized film mammograms," RSNA 2003, Chicago, 30 November–5 December 2003, p. 387.

Radiology

Heang-Ping Chan, PhD
Jun Wei, PhD
Berkman Sahiner, PhD
Elizabeth A. Rafferty, MD
Tao Wu, PhD
Marilyn A. Roubidoux, MD
Richard H. Moore, BA
Daniel B. Kopans, MD
Lubomir M. Hadjiiski, PhD
Mark A. Helvie, MD

Published online before print
10.1148/radiol.2373041657
Radiology 2005; 237:1075–1080

Abbreviations:
CAD = computer-aided detection
DBT = digital breast tomosynthesis
MQSA = Mammography Quality Standards Act
RBST = rubber-band straightening transform
ROC = receiver operating characteristic
3D = three-dimensional

From the Department of Radiology, University of Michigan, 1500 E Medical Center Dr, UHB1F510B, Ann Arbor, MI 48109-0030 (H.P.C., J.W., B.S., M.A.R., L.M.H., M.A.H.); and Department of Radiology, Massachusetts General Hospital, Boston, Mass (E.A.R., T.W., R.H.M., D.B.K.). Received September 27, 2004; revision requested December 2; revision received December 20; accepted January 12, 2005. Work conducted at the University of Michigan supported by USPHS grants CA91713 and CA95153 and U.S. Army Medical Research and Materiel Command grant DAMD17-02-1-0214. The development of the prototype digital breast tomosynthesis system and the clinical trial were supported by USAMRMC grant DAMD17-98-1-8309 awarded to Massachusetts General Hospital. **Address correspondence to** H.P.C. (e-mail: *chanhp@umich.edu*).

See Materials and Methods for pertinent disclosures.

The content of this paper does not necessarily reflect the position of the government, and no official endorsement of any equipment or product of any companies mentioned herein should be inferred.

Author contributions:
Guarantor of integrity of entire study, H.P.C.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, H.P.C., J.W.; clinical studies, E.A.R., T.W., R.H.M., D.B.K.; experimental studies, all authors; statistical analysis, J.W.; and manuscript editing, H.P.C., B.S., M.A.R., L.M.H., M.A.H.

Computer-aided Detection System for Breast Masses on Digital Tomosynthesis Mammograms: Preliminary Experience[1]

The purpose of the study was to design a computer-aided detection (CAD) system for breast mass detection on digital breast tomosynthesis (DBT) mammograms and to perform a preliminary evaluation of the performance of this system. Twenty-six patients were imaged with a prototype DBT system. Institutional review board approval and written informed patient consent were obtained. Use of the data set in this study was HIPAA compliant. The CAD system first screened the three-dimensional volume of the mass candidates by means of gradient-field analysis. Each mass candidate was segmented from the structured background, and its image features were extracted. A feature classifier was designed to differentiate true masses from normal tissues. The CAD system was trained and tested by using a leave-one-case-out method. The classifier calculated a mean area under the test receiver operating characteristic curve of 0.91 ± 0.03 (standard error of mean). The CAD system achieved a sensitivity of 85%, with 2.2 false-positive objects per case. The results demonstrate the feasibility of the authors' approach to the development of a CAD system for DBT mammography.

Mammography is considered the most cost-effective screening method for the early detection of breast cancer. However, the sensitivity of mammography is often limited by the presence of overlapping dense fibroglandular tissue in the breast. Dense parenchyma reduces the conspicuity of abnormalities and thus constitutes one of the main causes of missed breast cancer (1). The advent of full-field digital detectors offers opportunities to develop advanced techniques for improved imaging of dense breasts, such as digital tomosynthesis (2), stereomammography (3–7), and breast computed tomography (CT) (8). To our knowledge, these techniques are still under development and their potential influences on breast cancer detection remain to be investigated.

Digital tomosynthesis is based on the same principle as conventional tomography, which involves the use of a screen-film system as the image receptor for imaging body parts at selected depths. With conventional tomography, a series of projection exposures is accumulated on the same film when the x-ray source is moved about a fulcrum while the screen-film system is moved in the opposite direction. A drawback of conventional tomography is that each tomogram can depict only one plane at a selected depth with a relatively sharp focus. If the exact depth of interest is not known in advance or the abnormality encompasses a range of depths, then a tomogram at each depth will have to be acquired at separate imaging examinations, requiring additional radiation doses and examination time.

With digital tomosynthesis, the series of projection exposures is read out by the digital detector as separate projection views when the x-ray source moves to different locations about the fulcrum. Tomographic sections focused at any depth of the imaged volume can then be generated from the same series of projection images by using digital reconstruction techniques. Because of the wide dynamic range and the linear response of the digital detector, each projection image can

be acquired with a fraction of the x-ray exposure used to obtain a conventional projection radiograph. The total radiation dose required for digital tomosynthesis imaging may be kept at nearly the same as or only slightly higher than that required for conventional radiography. Properly designed digital reconstruction techniques have an additional advantage in that the depth resolution of tomosynthesis is generally much higher than that of conventional tomography. Thus, digital tomosynthesis makes it more practical to apply tomography to breast imaging in terms of radiation dose, examination time, and spatial resolution.

Digital breast tomosynthesis (DBT) mammography is one of the promising methods that may help reduce the camouflaging effects of dense breast tissue and improve the sensitivity of mammography for breast cancer detection in dense breasts. Several research groups are developing digital tomosynthesis methods for the reconstruction of tomographic sections from series of projection images (2,9–11). A study to compare DBT mammograms with conventional mammograms in breast cancer detection is underway (12).

Computer-aided detection (CAD) has been shown to improve breast cancer detection at mammography (13–15). Although the results of a preliminary evaluation indicated that breast lesions can be visualized more easily on DBT images than on conventional mammograms (12), to our knowledge, the overall detection sensitivity and specificity of DBT compared with those of conventional mammography remain to be investigated. With DBT, the number of reconstructed sections of each breast is very large. Even with 1-mm section thickness, the number of sections per breast will range from about 30 to more than 80. The time required to interpret a DBT case can be expected to be much longer than that required to interpret a conventional mammographic case.

With increases in radiologist workloads, the possibility of subtle lesions being overlooked may not be negligible. CAD will probably have a role in the reading of DBT mammograms, as it does in the reading of conventional mammograms. Thus, the purpose of our study was to design a CAD system for the detection of masses at DBT mammography and to perform a preliminary evaluation of the performance of this system.

Materials and Methods

Data Set

D.B.K. is the patent holder of the described DBT system. A data set of DBT cases was collected by the researchers (D.B.K., E.A.R., R.H.M., T.W.) at the Breast Imaging Research Laboratory of Massachusetts General Hospital with institutional review board approval. The recruited patients gave written informed consent. Use of the data set in this study was Health Insurance Portability and Accountability Act compliant. The patients were imaged with a prototype DBT system (GE Medical Systems, Milwaukee, Wis). This system has a flat-panel amorphous silicon detector with a pixel size of 0.1×0.1 mm. The DBT system acquired 11 projection-view mammograms of the compressed breast over a 50° arc in the mediolateral oblique view. The total radiation dose used to obtain the 11 projection-view mammograms was designed to be less than 1.5 times the dose used to obtain a single conventional (ie, screen-film) mammogram. DBT sections were reconstructed with 1-mm intersection spacing by using an iterative maximum-likelihood algorithm (9).

In this preliminary study, the DBT mammograms obtained in 26 patients aged 41–77 years (mean, 56 years; median, 56 years) were used. The number of DBT sections obtained per patient ranged from 37 to 89 (mean, 60.1), depending on the thickness of the compressed breast. Each patient case consisted of DBT sections of a single breast. The 26 cases included 23 breast masses and three areas of architectural breast distortion. Thirteen masses and two areas of architectural distortion were proved to be malignant at biopsy. Eight masses and the other area of architectural distortion were proved to be benign at biopsy. Two masses were determined to be benign by means of long-term follow-up or additional imaging. In each case, a Mammography Quality Standards Act (MQSA)-accredited radiologist (E.A.R.) with 5 years of experience in breast imaging determined the true location of the mass or area of architectural distortion on the basis of the diagnostic information. The longest diameters of the lesions ranged from 5.4 to 29.4 mm (mean, 14.2 mm; median, 12.1 mm), as estimated on the DBT section intersecting the lesion at approximately its largest cross section by an MQSA-accredited radiologist (M.A.H.) with 17 years of experience in breast imaging. The distribution of the longest diameters of the masses or areas of architectural distortion is shown in Figure 1. The distribution of breast density among the 26 breasts in terms of Breast Imaging Reporting and Data System category, as estimated by one of the MQSA-accredited radiologists (M.A.H.) by viewing the digitized screen-film mammograms, is shown in Figure 2.

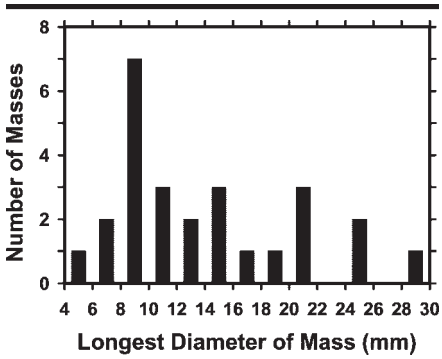


Figure 1. Distribution of longest diameters of the 23 masses and three areas of architectural distortion, as estimated on the DBT section intersecting the lesion at approximately its largest cross section.

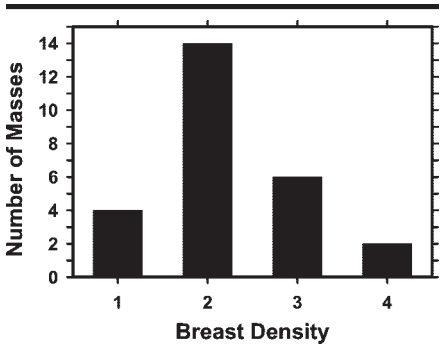
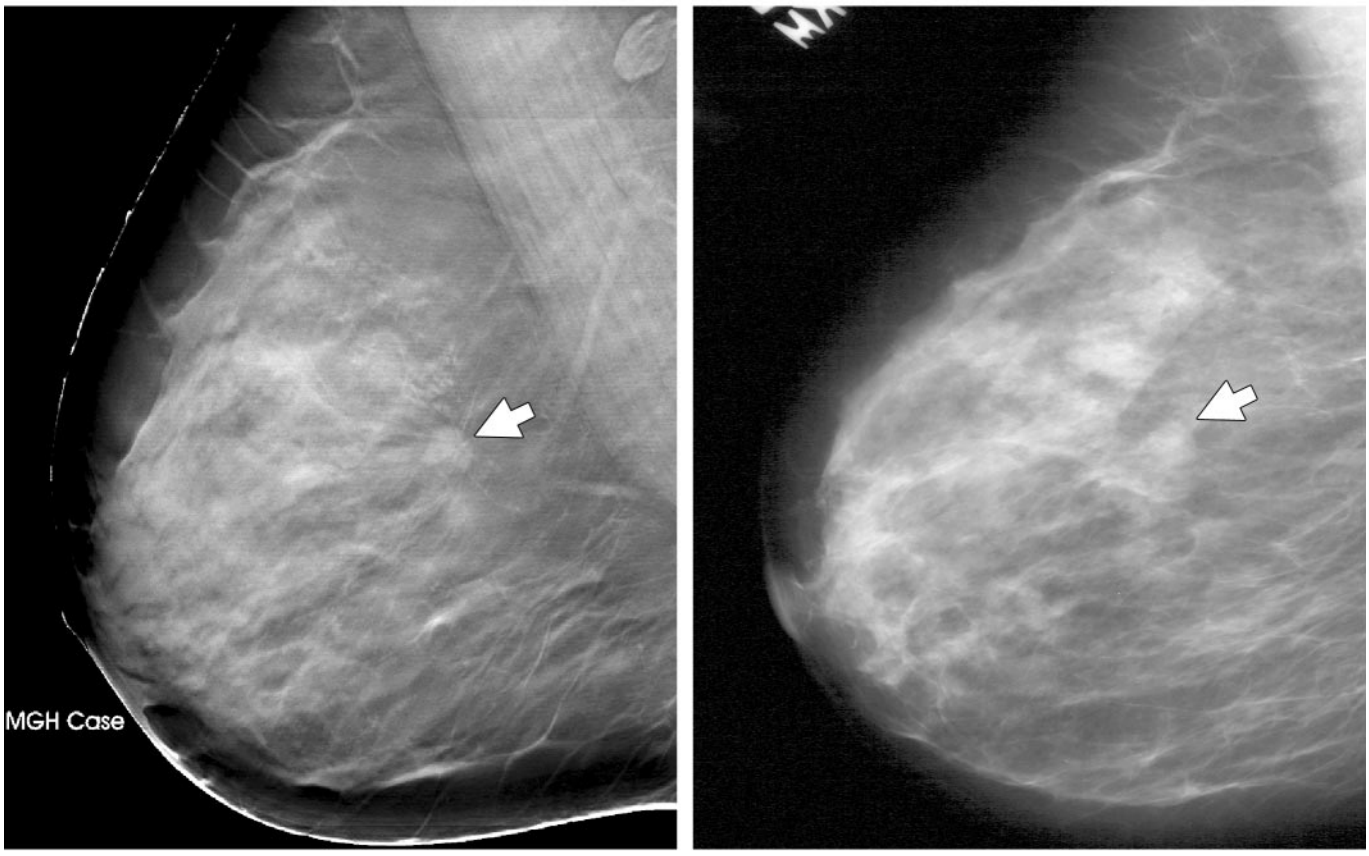


Figure 2. Distribution of breast density in terms of Breast Imaging Reporting and Data System category for the 26 breasts, as estimated from the conventional mammograms by an MQSA-accredited radiologist (M.A.H.).

An example of a DBT section intersecting a spiculated mass is shown in Figure 3a. For comparison, the same mass depicted in the same view on a conventional mammogram is shown in Figure 3b. The spicules of the mass are much more conspicuous on the DBT section than on the conventional mammogram, probably because of the reduced structured background on the DBT image.

Computerized Detection

The CAD mass detection system was developed in the CAD Research Laboratory at the University of Michigan. The



a. b.

Figure 3. Mass (arrow) depicted in the mediolateral oblique view on **(a)** DBT and **(b)** screen-film mammograms. The spicules of the mass are much more conspicuous in **a**.

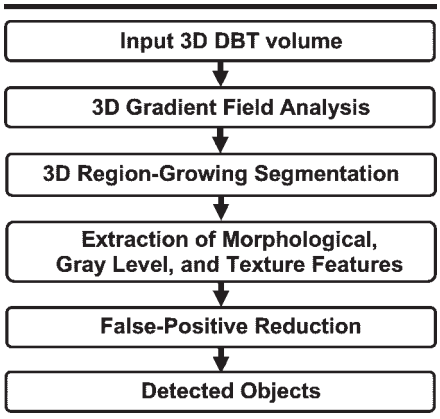


Figure 4. Schematic outline of CAD system steps for mass detection on DBT mammograms. 3D = three-dimensional.

system includes several major steps, including prescreening, segmentation, feature extraction, and false-positive object reduction, as shown in Figure 4. For a given case, the DBT section containing the entire breast volume was input into the CAD system for processing. The section thickness was linearly interpolated to 0.1 mm in the direction perpendicular to the detector plane so that the voxels in the data set were converted to 0.1 × 0.1 × 0.1-mm isotropic cubes.

In the prescreening step, three-dimensional (3D) gradient-field analysis of the volumetric data set in each case was performed to detect lesion candidates. To reduce noise in the gradient calculation, the image voxels were first averaged over every 2 × 2-voxel region to obtain a smoothed volumetric data set. The gradient-field analysis was performed in a spherical region that had a radius of about 6 mm and was centered at each voxel of the breast volume. The gradient vector at each smoothed voxel in the spherical region was computed, and the direction of the gradient vector was projected to the radial direction from the central voxel to the smoothed voxel. The average gradient direction over a spherical shell of voxels at a radius, *R(k),* of *k* voxels from the central voxel was calculated as the mean of the gradient directions over voxels on three adjacent spherical shells: $R(k-1)$, *R(k),* and $R(k+1)$. Finally, the gradient-field convergence at the central voxel was determined to be the maximum of the average gradient directions among all shells in the spherical region. Gradient-field convergence calculation was performed over all voxels in the breast region to result in a 3D gradient-field image.

The CAD algorithm then identified the locations of high-gradient convergence on the 3D gradient-field image as the locations of mass candidates. A 256 × 256 × 256-voxel volume of interest was centered at each location. The object in each volume of interest was segmented by using a 3D region-growing method with which the location of high-gradient convergence was used as the starting point and the object was allowed to "grow" across multiple sections. In this study, region growing was guided by the radial gradient magnitude. The growth of the object was terminated where the radial gradient reached a threshold value that was adaptively selected for the local object. After region growing, all con-

nected voxels constituting the object were labeled. The 3D object characteristics could then be extracted from the object.

Three groups of features—morphologic features, gray-level features, and texture features—were extracted from the segmented object. Morphologic feature descriptors included the volume in terms of the number of voxels in the object, the volume change before and after 3D morphologic opening by a spherical element with a 5-voxel radius, the surface area, the maximum perimeter of the segmented object among all sections intersecting the object, and the longest diameter of the object. The compactness of the object was described in terms of the percentage of overlap with a sphere of the same volume centered at the centroid of the object. The gray-level features included the contrast of the object relative to the surrounding background; the minimal and maximal gray levels; and the characteristics derived from the gray-level histogram of the object, such as skewness, kurtosis, energy, and entropy.

The texture features were described by using run-length statistics as follows: On each section, the cross section of the 3D object was treated as an object on a two-dimensional image. We applied the rubber-band straightening transform (RBST) that we previously developed for analysis of masses on two-dimensional mammograms (16) to the object. A 60-pixel-wide region around the object margin was transformed into a rectangular coordinate system. Sobel filtering in the x and y directions was then applied to the RBST image to generate gradient images in the two directions. A gradient-magnitude image of the transformed rectangular object margin was derived from these gradient images as the square root of the sum of the squares of the gradients at each corresponding pixel of these images.

Five run-length statistics texture features were extracted from the gradient-magnitude image in the horizontal and vertical directions: short-runs emphasis, long-runs emphasis, gray-level nonuniformity, run-length nonuniformity, and run percentage. A detailed description of the RBST and of the run-length statistics texture features for mammographic masses can be found in the literature (16,17). For a 3D object in the DBT data set, each run-length statistics texture feature was obtained by averaging the corresponding feature values over sections containing the segmented object.

Data and Statistical Analyses

Because of the limited data set available for this preliminary study, a leave-one-case-out resampling technique was used to train and test the performance of the CAD system. A classifier was trained to differentiate true masses from false-positive objects. The classifier was based on linear discriminant analysis and stepwise feature selection (18) that were designed with the training subset in each leave-one-case-out cycle. The trained classifier was applied to the lesion candidates in the left-out case such that each object was assigned a discriminant score. The test performance of the linear discriminant analysis classifier in differentiating true from false masses in the feature classification step of the CAD system was evaluated by performing receiver operating characteristic (ROC) analysis (19) of the discriminant scores of objects in the left-out cases. The area under the ROC curve and its standard error were obtained by using the ROCKIT program (version 9.1; Charles E. Metz, University of Chicago, Chicago, Ill), which uses maximum-likelihood estimation to fit a binormal ROC curve to the test discriminant scores output by the classifier.

Free-response ROC analysis was used to evaluate the test performance of the CAD system. A decision threshold was applied to the test discriminant score of each detected object. When an object had a discriminant score above the threshold, the location of that object was compared with the location of the true mass in that case. An object was considered to be true-positive if the centroid of the true mass marked by the radiologist was within the volume of the object; otherwise, the object was considered to be false-positive. For each decision threshold, the detection sensitivity and the average number of false-positive objects per case were determined on the basis of the entire data set. The free-response ROC curve was generated by varying the decision threshold over a range of values.

Results

Figure 5a and 5b shows an example of a section through a mass in a volume of interest and of the mass boundary determined by using 3D region-growing segmentation, respectively. An example of RBST applied to the section containing the mass and of the gradient image derived from the RBST image is shown in Figure 5c and 5d, respectively. The spicules radiating from the mass are approximately in the vertical direction, and the segmented boundary of the mass is transformed to a straight line, forming the upper edge of the rectangular RBST image.

To design the linear discriminant analysis classifier for false-positive object reduction, the stepwise feature selection procedure was used to select the most effective subset of features from the available feature pool and thus reduce the dimensionality of the feature space for the classifier (18,20). An average of seven features were selected from the available feature pool. The most often selected features included object contrast, minimal gray level, volume change before and after 3D morphologic opening, maximal perimeter, compactness, and two run-length statistics texture features—horizontal short-runs emphasis and gray-level nonuniformity. The ROC curve derived from the test discriminant scores of the masses and normal objects is shown in Figure 6. The area under the ROC curve reached 0.91 ± 0.03.

In the prescreening step, 100% of the masses and architectural distortions were detected, with an average of 29 false-positive objects per case. The overall test performance of the CAD system after false-positive object reduction is illustrated by the free-response ROC curve shown in Figure 7. The system achieved sensitivities of 85% (22/26) with 2.2 false-positive objects per case and 80% (21/26) with 2.0 false-positive objects per case in this preliminary study.

Discussion

In this preliminary study, we used a 3D approach that takes advantage of the volumetric nature of tomosynthesis reconstruction. Prescreening of lesion candidates, image segmentation, and feature extraction were performed in the volumetric data set for each breast. The prescreening and segmentation methods developed for 3D processing are effective for locating true lesions. Although the training samples in this study were small, the overall performance of the system is promising. Therefore, the results of this study demonstrate the feasibility of our approach to the development of a CAD system for assisting radiologists in detecting masses on DBT mammograms. Further improvement in the performance of the system can be expected with use of a larger data set for training the algorithms.

With DBT mammography, the struc-

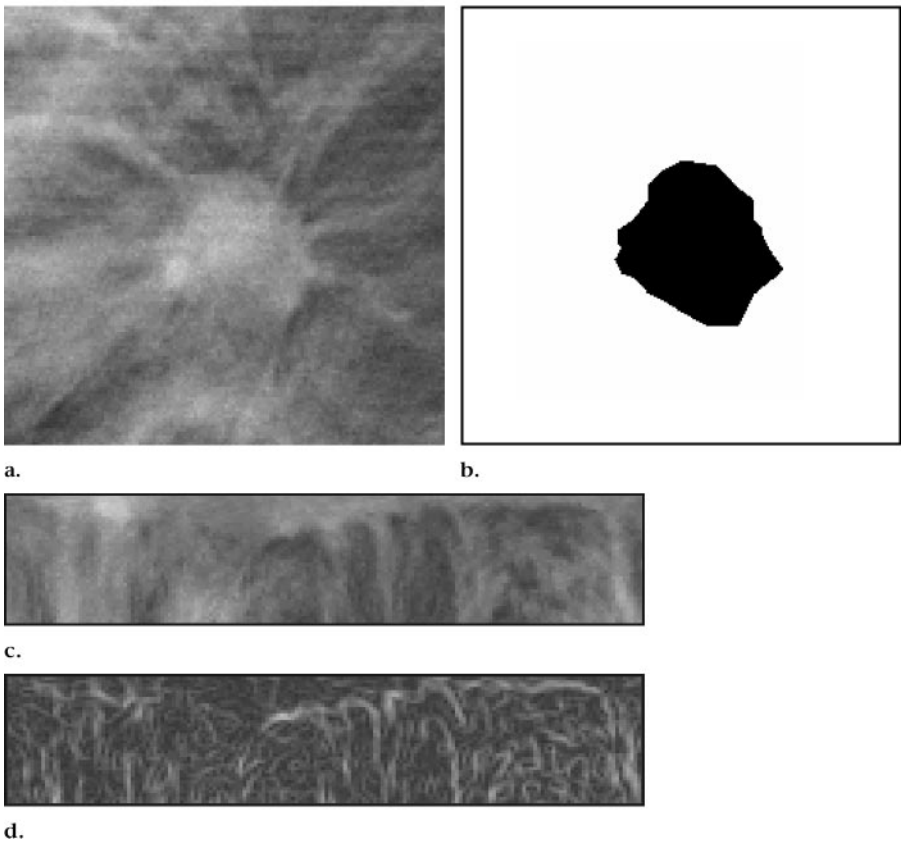


Figure 5. **(a)** DBT mammographic section intersecting a spiculated breast mass. **(b)** Mass in **a** after 3D region-growing segmentation. **(c)** RBST image of a 60-pixel-wide region around the same mass. The segmented mass boundary is transformed into a straight line, forming the upper boundary of the rectangular RBST image. **(d)** Gradient-magnitude image derived from Sobel filtering of the RBST image in **c**.

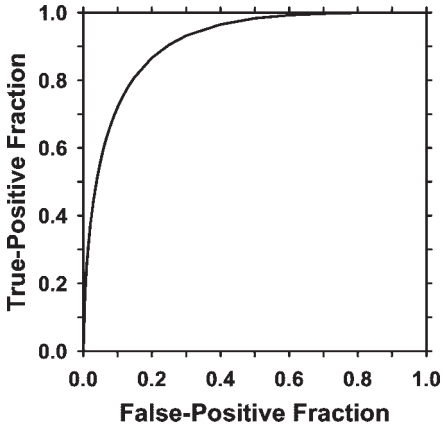


Figure 6. ROC curve showing the performance of the linear discriminant classifier obtained from leave-one-case-out testing. The area under the ROC curve was 0.91 ± 0.03, indicating that the classifier was effective in reducing the number of false-positive objects.

tured background such as the dense fibroglandular tissue was suppressed on the reconstructed DBT sections. However, DBT is different from CT in that the overlapping tissues are reduced but not totally eliminated. Tomosynthesis reconstruction left residual overlapping tissue on the DBT sections. Similarly, the shadow of a lesion can be seen on most DBT sections, even though the actual size of the mass may be only a fraction of the breast thickness. In addition, the voxel dimension in the z direction (ie, the direction perpendicular to the sections) on the reconstructed sections is 10 times larger than that in the x-y plane (ie, the planes of the sections). Therefore, the boundary of an object in the z direction is not as well defined as that in the x-y plane.

The features extracted in three dimensions may have a strong directional dependence. For example, in this study we extracted texture features along the x-y plane only, and a 3D texture feature was obtained by averaging the corresponding two-dimensional texture values over sections containing the object. For true 3D texture analysis, the texture feature values should be calculated in the shell of voxels surrounding the object or on the planes that intersect the object centroid from different directions. We will investigate the potential directional effects of the features on false-positive object reduction when a larger data set becomes available.

A limitation caused by the small data set in this study is the possibility that the distributions of the characteristics of the masses and the breast parenchyma in this data set were not statistically similar to those in the patient population. Although the results appear to be promising, the methods and features used may have been biased toward the specific data set used. Further studies are needed to evaluate the robustness of these computer vision techniques in a larger data set.

For DBT imaging, the raw data were acquired as 11 projection-view mammograms. On average, each projection-view mammogram was obtained by using about 14% of the radiation dose used to obtain a conventional mammogram. A projection-view mammogram is therefore noisier than a conventional mammogram. However, the 11 projection-view mammograms offer the advantage that a lesion will be projected at slightly different angles, and, thus, there will be somewhat different overlapping tissues on each view. A lesion that may be camouflaged by dense tissue on some views may become more conspicuous on other views. In addition, overlapping tissues that mimic lesions on some views may mimic lesions to a lesser degree on other views. If a CAD lesion detection system is applied to projection-view mammograms, the complementary information derived from the different projection-view mammograms may be used to improve sensitivity and reduce the number of false-positive objects. We are studying the feasibility of developing a CAD system for detecting lesions on projection-view mammograms and investigating methods to merge the information from the 11 projection-view mammograms. In future studies, this approach will be compared with the current approach of detecting lesions on reconstructed DBT volumetric data sets.

Furthermore, although with our current lesion-detection algorithm, DBT sections reconstructed from the iterative maximum-likelihood algorithm are used as input, we expect that our image-processing methods will not strongly depend on the reconstruction method for

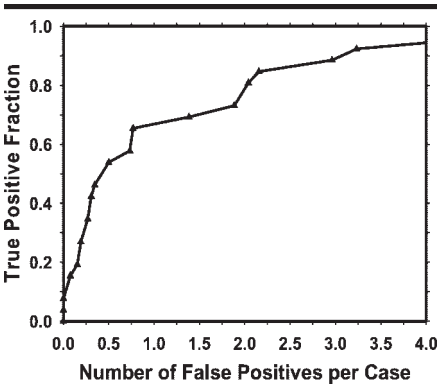


Figure 7. Free-response ROC curve for the test performance of the CAD system for DBT mammography. The current CAD system achieved 85% sensitivity, with 2.2 false-positive objects per case.

generating the DBT sections as long as the image quality of the reconstructed sections is reasonable. The effects of the factors that may affect image quality—including image acquisition technique, number of projection views, tomographic angle, reconstruction method, and section thickness—on lesion detection accuracy will have to be investigated when DBT cases obtained with different methods and parameters become available in the future.

Acknowledgment: The authors are grateful to Charles E. Metz, PhD, for the ROCKIT program provided from the University of Chicago Web site *(http://xray.bsd.uchicago.edu/krl/index.htm)*.

References

1. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992;184:613–617.
2. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. Radiology 1997;205:399–406.
3. Goodsitt MM, Chan HP, Hadjiiski LM. Stereomammography: evaluation of depth perception using a virtual 3D cursor. Med Phys 2000;27:1305–1310.
4. Goodsitt MM, Chan HP, Darner KL, Hadjiiski LM. The effects of stereo shift angle, geometric magnification and display zoom on depth measurements in digital stereomammography. Med Phys 2002;29:2725–2734.
5. Getty DJ, Pickett RM, D'Orsi CJ. Stereoscopic digital mammography: improving detection and diagnosis of breast cancer. Berlin, Germany: International Congress Series 2001;1230:506–511.
6. Chan HP, Goodsitt MM, Hadjiiski LM, et al. ROC study comparing radiologists' performances in evaluating breast lesions on stereoscopic and single-projection digital specimen mammograms (abstr). Med Phys 2003;30:1456.
7. Chan HP, Goodsitt MM, Hadjiiski LM, et al. Effects of magnification and zooming on depth perception in digital stereomammography: an observer performance study. Phys Med Biol 2003;48:3721–3734.
8. Boone JM, Nelson TR, Lindfors KK, Seibert JA. Dedicated breast CT: radiation dose and image quality evaluation. Radiology 2001;221:657–667.
9. Wu T, Stewart A, Stanton M, et al. Tomographic mammography using a limited number of low-dose cone-beam projection images. Med Phys 2003;30:365–380.
10. Suryanarayanan S, Karellas A, Vedantham S, et al. Evaluation of linear and nonlinear tomosynthetic reconstruction methods in digital mammography. Acad Radiol 2001;8:219–224.
11. Dobbins JT, Godfrey DJ. Digital x-ray tomosynthesis: current state of the art and clinical potential. Phys Med Biol 2003;48:R65–R106.
12. Rafferty EA, Georgian-Smith D, Kopans DB, Hall DA, Moore R, Wu T. Comparison of full-field digital tomosynthesis with two view conventional film screen mammography in the prediction of lesion malignancy (abstr). Radiology 2002;225(P):268.
13. Chan HP, Doi K, Vyborny CJ, et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computer-aided diagnosis. Invest Radiol 1990;25:1102–1110.
14. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12 860 patients in a community breast center. Radiology 2001;220:781–786.
15. Helvie MA, Hadjiiski LM, Makariou E, et al. Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection: a pilot clinical trial. Radiology 2004;231:208–214.
16. Sahiner B, Chan HP, Petrick N, Helvie MA, Goodsitt MM. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Med Phys 1998;25:516–526.
17. Galloway MM. Texture classification using gray level run lengths. Comput Graph Image Process 1975;4:172–179.
18. Sahiner B, Chan HP, Petrick N, Wagner RF, Hadjiiski LM. Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size. Med Phys 2000;27:1509–1522.
19. Metz CE, Herman BA, Shen JH. Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously distributed data. Stat Med 1998;17:1033–1053.
20. Chan HP, Sahiner B, Wagner RF, Petrick N. Classifier design for computer-aided diagnosis: effects of finite sample size on the mean performance of classical and neural network classifiers. Med Phys 1999;26:2654–2668.

Computerized pectoral muscle identification on MLO-view mammograms for CAD applications

Chuan Zhou, Lubomir M. Hadjiiski, Chintana Paramagul, Berkman Sahiner, Heang-Ping Chan, Jun Wei
Department of Radiology, The University of Michigan, Ann Arbor, MI 48109-0904

ABSTRACT

Automatic identification of the pectoral muscle on MLO view is an essential step for computerized analysis of mammograms. It can reduce the bias of mammographic density estimation, will enable region-specific processing in lesion detection programs, and also may be used as a reference in image registration algorithms. We are developing a computerized method for the identification of pectoral muscle on mammograms. The upper portion of the pectoral edges was first detected to estimate the direction of the pectoral muscle boundary. A gradient-based directional (GD) filter was used to enhance the linear texture structures, and then a gradient-based texture analysis was designed to extract a texture orientation image that represented the dominant texture orientation at each pixel. The texture orientation image was enhanced by a second GD filter. An edge flow propagation method was developed to extract edges around the pectoral boundary using geometric features and anatomic constraints. The pectoral boundary was finally generated by a second-order curve fitting. 118 MLO view mammograms were used in this study. The pectoral muscle boundary identified on each image by an experienced radiologist was used as the gold standard. The accuracy of pectoral boundary detection was evaluated by two performance metrics. One is the overlap percentage between the computer-identified area and the gold standard, and the other is the root-mean-square (RMS) distance between the computer and manually identified pectoral boundary. For 118 MLO view mammograms, 99.2% (117/118) of the pectoral muscles could be identified. The average of the overlap percentage is 94.8% with a standard deviation of 20.9%, and the average of the RMS distance is 4.3 mm with a standard deviation of 5.9 mm. These results indicate that the pectoral muscle on mammograms can be detected accurately by our automated method.

Keywords: Computer-aided detection, Pectoral muscle trimming, Breast density estimation, Directional gradient filter

1. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women[1, 2]. At present, the most successful method for the early detection of breast cancer is screening mammography[3]. It has been demonstrated that an effective computer-aided diagnosis (CAD) system can provide a second opinion to the radiologists and improve the accuracy of detection and characterization of mammographic abnormalities, which, in turn, may reduce unnecessary biopsies. Studies have shown that there is a strong positive correlation between breast parenchymal density on mammograms and breast cancer risk.[1, 4-6] The relative risk is estimated to be about 4-6 times higher for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% of parenchymal densities. Mammograms are analyzed visually by radiologists, the qualitative response may vary from radiologist to radiologist due to the subjective nature of visual analysis. We have previously developed a computerized system, mammographic density estimator (MDEST), to estimate breast density automatically on digitized film mammograms.[7] For each mammogram, the breast region was first segmented by breast boundary detection and, for the mediolateral oblique (MLO) view, with additional pectoral muscle trimming. A gray level threshold was then automatically determined to segment the dense tissue from the breast region. The breast density was estimated as the percentage of the segmented dense area relative to the breast area. Our preliminary study indicated that the computer-estimated mammographic breast density correlated closely with the "reference standard" obtained by averaging five experienced radiologists' manual segmentations and the average bias was much less than that of the radiologists' visual estimation.

Automatic identification of the pectoral muscle is an essential step for computerized analysis of mammograms. Accurate segmentation of the pectoral muscle on MLO-view mammograms can reduce the bias of mammographic

Medical Imaging 2005: Image Processing, edited by J. Michael Fitzpatrick, Joseph M. Reinhardt, Proc. of SPIE Vol. 5747 (SPIE, Bellingham, WA, 2005)
1605-7422/05/$15 · doi: 10.1117/12.595693

density estimation and improve the performance of our MDEST method. It will enable region-specific processing in lesion detection programs to reduce false negatives. False positives can be reduced if the detected objects in the pectoral muscle area can be selectively suppressed. The identification of the pectoral muscle may also be used as a reference in image registration algorithm for multiple-view analysis of mammograms.

In our preliminary study [7], the pectoral muscle was trimmed using a gradient-based pectoral edge detection method: the initial edge in the pectoral region was first found as the maximum gradient point by a line-by-line gradient analysis from the chest wall to the breast boundary. An edge validation process was then performed to remove the false pectoral muscle edges using a line fitting method, and a coarse direction of the pectoral edges was estimated from the validated edges. The remaining pectoral edges were extrapolated along the estimated pectoral direction. Finally, a second order curve was fitted to the detected pectoral edges to generate the pectoral boundary. Using the above method, 74.6% of the pectoral muscles were determined by visual judgment to be correctly identified in this preliminary study.

The purpose of this study is to improve the performance of our previously developed pectoral muscle segmentation method. Accurate identification of the pectoral muscle on mammograms is challenging, especially for the improperly positioned MLO-view images and the images containing dense glandular tissues overlapping with the pectoral muscle region. In this work, we developed a two-stage gradient-based texture analysis method to detect the pectoral boundary. In the first stage, linear texture structures were enhanced and the directional gradients were computed using a directional filter. In the second stage, a texture orientation image was derived as the dominant texture orientation at each pixel. A diffusion filter was used to estimate the global direction of the pectoral boundary. An edge flow propagation method was developed to extract the pectoral edges with the guidance of the estimated global direction.

2. MATERIALS AND METHODS

2.1 Materials

In this study, 118 MLO-view mammograms from 103 patients were randomly selected from the patient files in the Radiology Department at the University of Michigan. Data collection was approved by the Institutional Review Board and individual patient informed consent was waived. The mammograms were acquired with Mammography Quality Standards Act (MQSA) approved GE DMR (Milwaukee, Wisconsin) mammography units using Kodak MR2000 screen/film systems. All films were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50 μm×50 μm and 4096 gray levels. The resolution of the mammograms was reduced to 800 μm × 800 μm for segmentation of the pectoral muscle.

2.2 Pectoral muscle identification

Figure 1 summarizes the automatic pectoral muscle identification scheme. The interference due to overlapping of the glandular tissue on the pectoral muscle region is first reduced by smoothing the mammogram using an edge preserving anisotropic diffusion filter [8]. Because less glandular tissue appears at the upper region of the pectoral muscle, the upper portion of the pectoral boundary usually remains sharp after smoothing and can be detected robustly by searching the maximum horizontal gradients on the diffused image. The extrapolation of the detected upper pectoral boundary provides a coarse global direction of the pectoral boundary. To refine the entire pectoral boundary, a gradient-based directional (GD) filter was first employed to enhance the linear texture structures on the mammogram. The orientation of the digitized image could be automatically determined by the curvature of the breast boundary. For example, if the image was positioned such that the chest wall was on the right side, it could be assumed that the pectoral boundary is at a direction approximately from the top-left to the bottom-right with less than 45 degree deviation. Therefore, in our study, the kernel of the GD filter was designed as a step function with 45 degree orientation. After the pectoral edge was enhanced by the GD filter, a gradient-based texture analysis[9] was used to compute an orientation image which represented the dominant texture orientation at each pixel. The orientation image was smoothed using an edge preserving mean shift algorithm[10] that iteratively shifted each pixel to the average of the pixels in its neighborhood. The texture patterns with dominant texture orientations directing from the top-left to the bottom-right, which were more likely to be the pectoral edges, were enhanced by applying a second GD filter to the smoothed orientation image. Candidate edges of the pectoral muscle were detected on the enhanced orientation image using a ridge-tracking

algorithm. The ridges were tracked by searching for the local maximum along the coarse global direction estimated, as described above, by the upper pectoral boundary on the anisotropic diffused image. With the guidance of the estimated global direction of the pectoral boundary and the anatomical constraints, an edge flow propagation algorithm was then used to extract the boundary points of the pectoral muscle by pruning the edges that are less likely to lie on the pectoral boundary. A second order curve fitting was finally used to generate the pectoral muscle boundary. Figure 2 shows examples of the intermediate images of pectoral boundary enhancement and edge tracking corresponding to the various stages shown in the flowchart in Figure 1.

3. RESULTS

An experienced MQSA-radiologist used a graphical user interface to manually draw the pectoral muscle boundary on each MLO-view mammogram, which was then used as the gold standard for the evaluation of the performance of our pectoral muscle detection program.

For each MLO view mammogram, the accuracy of pectoral boundary detection was evaluated by two performance metrics: the percentage of overlap, defined as the ratio of the overlap area between the computer detected pectoral muscle area and the gold standard relative to the gold standard, and the root-mean-square (RMS) distance obtained by calculating the shortest distance point by point between the computer-identified pectoral boundary and the manually marked pectoral boundary. For the data set of 118 MLO view mammograms, 99.2% (117/118) of the pectoral muscles could be identified, the average of the percent overlap area is 94.8% with a standard deviation of 20.9%, the average of the RMS distance is 4.3 mm with a standard deviation of 5.9 mm.

Figure 3 shows some examples of pectoral boundary identification on mammograms. The computer identified pectoral boundaries were shown in white lines and the dark lines show the radiologist's hand drawn boundaries. Figure 3 (a)-(b) show the pectoral boundary can be identified accurately on mammograms with weak pectoral edges (figure 3(a)) and a large area of dense tissues overlapping on the pectoral muscle area (shown in figure 3(b)). Figure 3(c)-(d) show two examples of less accurate pectoral boundaries detected by the computer. Figure 3(e) shows the only case in this data set that the computer failed to detect the boundary.

4. CONCLUSION

The newly developed gradient-based directional filter and the dominant texture orientation estimation method can enhance the pectoral boundary regions. The edge flow propagation method can accurately extract pectoral edges to generate the pectoral boundary. Automatic pectoral muscle identification will provide the foundation for many mammographic image analysis tasks in CAD applications.

5. ACKNOWLEDGMENT

This work is supported by USPHS grant CA 95153 and U.S Army Medical Research and Materiel Command grant DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

REFERENCES

1. S. H. Landis, T. Murray, S. Bolden and P. A. Wingo, "Cancer statistics, 1998," *CA Cancer J Clin* **48**, 6-29, 1998.
2. C. Byrne, C. R. Smart, C. Cherk and W. H. Hartmann, "Survival advantage differences by age: Evaluation of the extended follow-up of the breast cancer detection demonstration project," *Cancer* **74**, 301-310, 1994.
3. H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," *Breast cancer, diagnosis and treatment*, I. M. Ariel and J. B. Cleary, Eds., 152-172, McGraw-Hill, 1987.

4. J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *Am. J. Roentgenol.* **126**, 1130-1139, 1976.
5. J. N. Wolfe, A. F. Saftlas and M. Salene, "Evaluation of mammographic densities: A case-control study," *AJR* **148**, 1087-1092, 1987.
6. A. F. Saftlas, R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane and J. N. Wolfe, "Mammographic densities and risk of breast cancer," *Cancer* **67**, 2833-2838, 1991.
7. C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner and L. M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Medical Physics* **28**, 1056-1069, 2001.
8. P. Perona and J. Malik, " Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern analysis and machine intelligence* **12**, 629-639, 1990.
9. A. R. Rao and B.G.Schunck, "Computing oriented texture fields," *CVGIP: Graphical models image processing* **53**, 157-185, 1991.
10. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. on Pattern analysis and machine intelligence* **24**, 603-619, 2003.

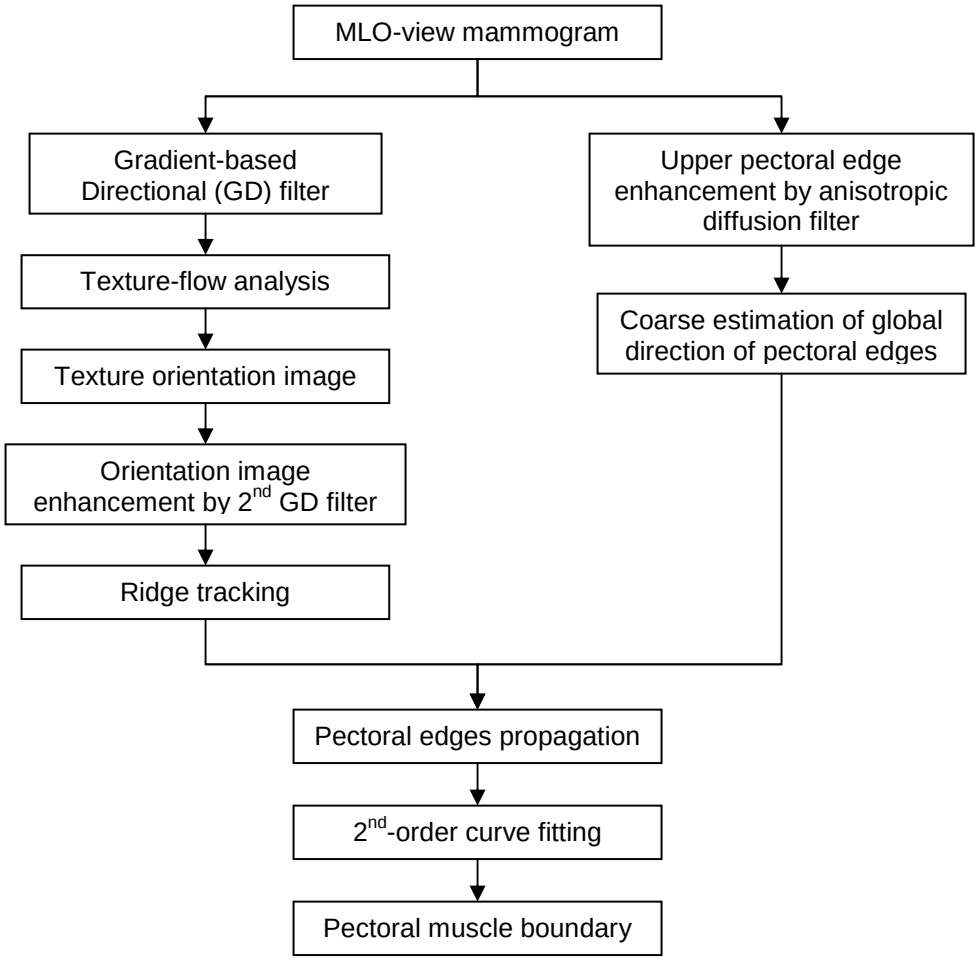


Figure 1. Automated pectoral muscle detection scheme

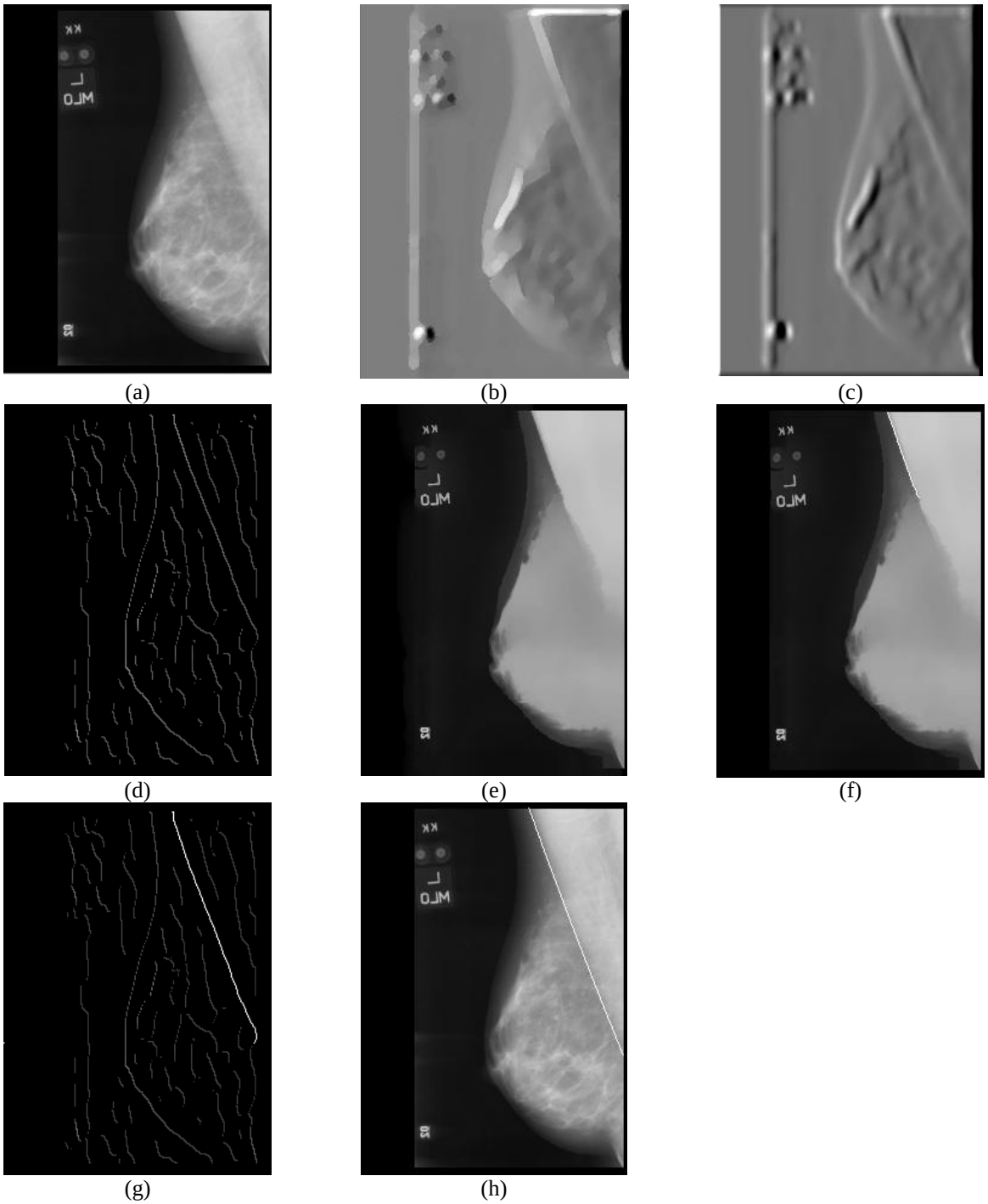


Figure 2. Example of boundary enhancement and segmentation of pectoral muscle. (a) original image; (b) texture orientation image after first GD filter and texture-flow analysis; (c) ridge image enhanced by the 2nd GD filter; (d) tracked ridges; (e) smoothed image using anisotropic diffusion filter; (f) initial pectoral edges detected from the smoothed image in (e) for the estimation of the coarse direction of the pectoral boundary; (g) propagated pectoral edges on the ridge image (c) with the guidance of the coarse direction estimated from the smoothed image shown in (f); (h) the final identified pectoral boundary after 2nd order curve fitting.

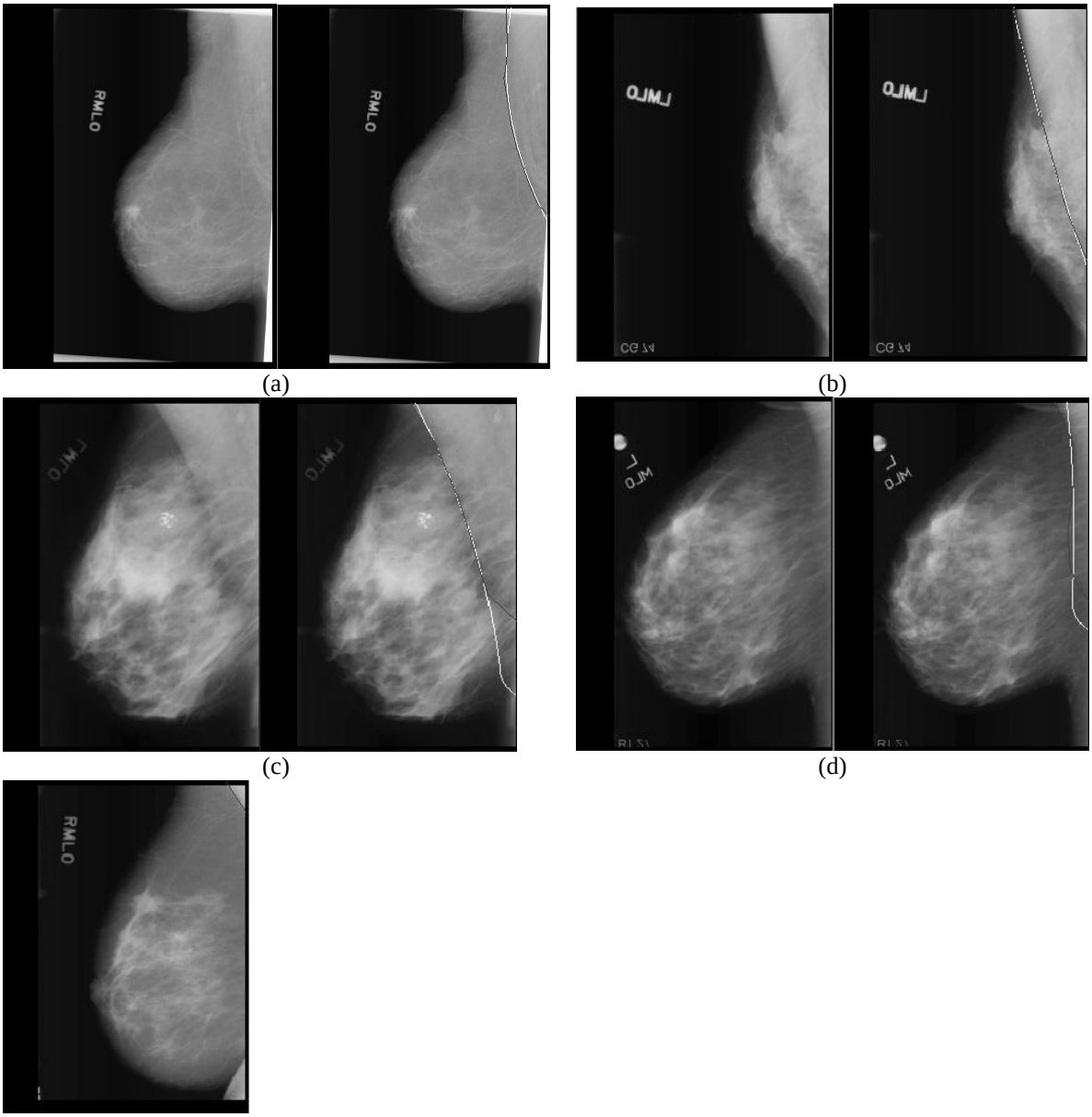


Figure 3. Examples of pectoral boundary segmentation on mammograms. (a)-(b): accurate identification of pectoral boundary; (c)-(d): less accurate identification of pectoral boundary; (e) the only mammogram in our data set that the computer failed to identify the pectoral muscle due to the small portion of the pectoral muscle area within the breast region.

Effects of the Continuous and Discrete Confidence Rating Scales in ROC Observer Studies

Lubomir Hadjiiski*, Heang-Ping Chan, Berkman Sahiner, Mark A. Helvie, Marilyn A. Roubidoux

Department of Radiology, University of Michigan, Ann Arbor, MI 48109

ABSTRACT

We previously conducted an observer study evaluating radiologists' performance for characterization of mammographic masses on serial mammograms with and without CAD. 253 temporal image pairs (138 malignant and 115 benign) from 96 patients containing masses on serial mammograms were used. The interval change characteristics of the masses on each temporal pair were analyzed by our CAD program to differentiate malignant and benign masses. The classifier achieved a test A_z value of 0.87 for the data set. Eight MQSA radiologists and 2 fellows assessed the temporal masses and provided estimates of the likelihood of malignancy (LM) and BI-RADS assessment without and then with CAD. The LM estimates were provided on a quasi-continuous confidence-rating scale (CRS) of 1 to 100. In the current study we investigated the effects of using discrete CRS with fewer categories on ROC analysis. We simulated three discrete CRSs containing 5, 10, and 20 categories by binning the radiologists' LM quasi-continuous ratings. For the ten radiologists, without CAD, the average A_z in estimating the LM for the 5, 10, 20 and 100 category CRSs were 0.788, 0.786, 0.785, and 0.787, respectively. With CAD, the observers' A_z improved to 0.845, 0.843, 0.844, and 0.843, respectively. The improvement was statistically significant ($p<0.011$) for each CRS. The partial area index for the four CRSs without CAD was 0.198, 0.204, 0.200, and 0.206, respectively. With CAD the partial area index was also significantly improved to 0.369, 0.365, 0.369, and 0.366, respectively ($p<0.006$ for all CRSs). The use of continuous and discrete confidence-rating scales in this study had minimal effect on the analysis of observer performance.

Keywords: Computer-Aided Diagnosis, Continuous and Discrete Confidence Rating Scales, Interval Changes, ROC Observer Study, Classification, Mammography, Malignancy.

1. INTRODUCTION

The effect of the use of quasi-continuous or discrete confidence rating scales on receiver operating characteristic (ROC) observer study results has been studied by a number of researchers. Rockette et al[1] carried out an observer experiment using both five-point discrete scale and a quasi-continuous 100-point scale. The results of ROC analysis showed no statistically significant difference between the performance index A_z achieved with the two scales. However, they suggested that the use of quasi-continuous scale can be more reliable for ROC analysis because it can avoid the problem of "degenerate" data sets. King et al[2] performed an observer study to estimate the likelihood of the presence of abnormality on chest images using a quasi-continuous scale. Then they mapped the quasi-continuous observer ratings to a 5-point rating scale using two different sets of criteria for determining the range of each category and used ROC methodology to analyze the results. They concluded that the diagnostic accuracy derived from the quasi-continuous rating data are insensitive to the particular way those data are mapped to discrete categories. They also suggested that the use of a quasi-continuous scale is better in observer studies because of the insensitivity of the mapping to discrete categories and the reduced likelihood of "degenerate" data. Wagner et al[3] performed a Monte Carlo simulation study of multiple-reader, multiple-case ROC experiments to evaluate the data quantization effects. They concluded that the discretization in five categories can reduce the precision of ROC measurements, in comparison to that obtained from continues scale. Berbaum et al[4] suggested that quasi-continuous 101-point scale ratings fitted with a standard binormal model may sometimes yield inappropriate chance line crossings, reducing the statistical power to

* L. H. (correspondence): e-mail: lhadjiski@umich.edu

Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment,
edited by Miguel P. Eckstein, Yulei Jiang, Proceedings of SPIE Vol. 5749
(SPIE, Bellingham, WA, 2005) · 1605-7422/05/$15 · doi: 10.1117/12.595725

detect the differences between two experimental conditions. They concluded that the use of proper ROC models with the discrete confidence rating data may present better results, however, they stressed that this should be investigated further.

We have previously studied radiologists' performance of characterizing malignant and benign masses in single-view serial mammograms with and without CAD[5,6] using ROC methodology. The observers' estimate of the likelihood of malignancy of the lesions was collected on a quasi-continuous 100-point confidence-rating scale. We observed a statistically significant improvement (p=0.005) in the radiologists' performance when they used CAD compared to their performance without CAD. In this study, we examined the effects of the number of confidence ratings used in an ROC experiment on the results of ROC analysis. The observer rating data collected from the CAD mass characterization experiment were used as an example. We simulated the use of discrete confidence-rating scales with a small number of categories and compared the performance indices and statistical significance obtained with ROC analysis for the different confidence-rating scales.

2. MATERIALS AND METHODS

We previously conducted an observer ROC study evaluating radiologists' performance for characterization of mammographic masses on serial mammograms with and without CAD[6]. A brief description of the database used and the observer study design is given below.

2.1 Data set

Two hundred fifty three temporal image pairs (138 malignant and 115 benign) from 96 patients containing masses on serial mammograms were used. The mammograms in the database were collected from the patients who had undergone breast biopsy in our department. The selection criterion used in this study was that the case had serial exams in which a corresponding mass could be identified. The mammograms thus contained masses covering a range of sizes and conspicuity that are seen in clinical practice. Since all cases eventually underwent biopsy, interval change was observed for most of the masses even if they were found to be benign after biopsy. Thirty-four additional temporal pairs containing corresponding normal structures in the serial mammograms were also included. In this way the radiologist also had to distinguish mass-mimicking fibroglandular tissue from malignant masses, thus simulating a more realistic clinical situation. The temporal pairs had a time interval of 6 to 48 months. More than 55% of the pairs had a time interval of 12 months.

The mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of $50\mu m \times 50\mu m$ and 4096 gray levels. The image matrix size was reduced by averaging every 2 x 2 adjacent pixels and down-sampled by a factor of 2 to obtain images with a pixel size of $100\mu m \times 100\mu m$ for analysis by the computer. The interval change of a mass on a corresponding temporal pair was analyzed by the CAD system developed in our laboratory.

2.2 Classification of masses in serial mammograms

We have previously developed a novel classification technique that utilizes the current and prior information on serial mammograms to characterize the masses on corresponding mammographic views. The classification technique has been described in detail elsewhere[6,7]. The classifier was based on texture, morphological, and different features extracted from current and prior ROIs. The classifier was trained and tested using a leave-one-case-out resampling scheme and it achieved a test A_z value of 0.87 for the data set.

2.3 Observer ROC study

The observer study was designed to compare radiologists' performance on the classification of malignant and benign breast masses with and without CAD on single-view temporal pairs of mass ROIs. The ROIs extracted from the current and the prior mammograms containing the corresponding mass were displayed side-by-side on a display monitor. The observers' performance was evaluated under two reading conditions – reading with and without CAD[6]. In the first reading condition, the radiologist read the temporal image pair of the mass without computer aid. In the second

reading condition, the radiologist read the temporal pair with the computer classifier's relative malignancy rating of the mass displayed on the screen. The observer was asked to provide an estimate of the likelihood of malignancy and BI-RADS assessment of the mass under each reading condition (Fig. 1). The likelihood of malignancy estimates were provided on a quasi-continuous confidence-rating scale of 1 to 100 (1=benign, 100=high likelihood of malignancy). Eight MQSA radiologists and 2 fellows participated as observers.

A counter-balanced design was used in arranging the reading orders in different modes and the case orders in different reading sessions for the observers. This approach would minimize the potential effects such as learning, fatigue, and memorization on the outcomes of the observer experiments. A graphic user interface (GUI) was developed for the purpose of presenting the temporal pairs of mass ROIs to the radiologists and recording their ratings. Each observer underwent a training session before the actual reading sessions to familiarize them with the performance of the CAD system and the experimental procedure.

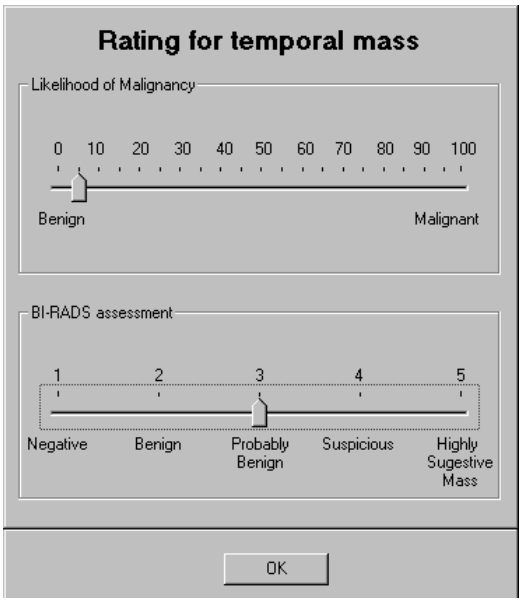


Figure 1. The GUI for collection of the likelihood of malignancy and BI-RADS ratings in our ROC study.

2.4 Quasi-continuous and discrete rating experiments

In the current study, the radiologists' quasi-continuous ratings on likelihood of malignancy without and with CAD were mapped to three discrete confidence-rating scales to simulate ROC experiments with fewer number of rating categories. A simple mapping by grouping every k adjacent ratings was chosen in this study (Fig. 2). We used three groupings of the adjacent ratings with k=20, k=10, and k=5, which resulted in three discrete rating scales, 5-point, 10-point, and 20-point, respectively. Based on the original quasi-continuous rating scale and the three simulated discrete rating scales, we studied the change in the observer performance accuracy and the change in the statistical significance of the results.

2.5 ROC analysis

The radiologists' classification accuracy based on the different confidence-rating scales was analyzed with ROC methodology. Their performances were quantified by the total area under the ROC curve, A_z, as well as the partial area index [8] calculated above a sensitivity threshold of 0.9, $A_z^{(0.90)}$. The area under the ROC curve was estimated

by the Dorfman-Berbaum-Metz (DBM) multi-reader multi-case (MRMC) methodology [9], in which the ROC curve was derived from binormal distributions fitted to the observer ratings by maximum likelihood estimation. The statistical significance of the difference in A_z between the different reading conditions for the different confidence-rating scales was also estimated by the DBM analysis.

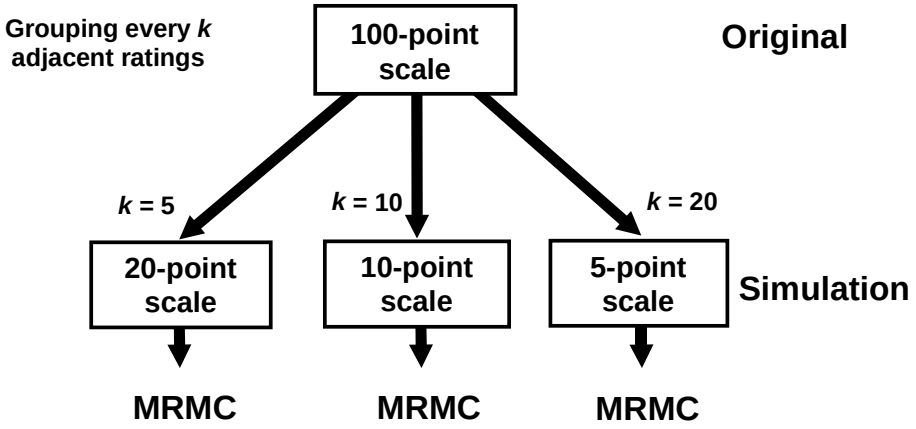


Figure 2. The block diagram of mapping the quasi-continuous confidence rating scale to the simulated discrete confidence rating scales.

3. RESULTS

For the ten radiologists, without CAD, the average A_z in estimating the likelihood of malignancy for the 5, 10, 20 and 100 category confidence rating scales was 0.788, 0.786, 0.785, and 0.787, respectively. The observers' A_z improved to 0.845, 0.843, 0.844, and 0.843, respectively with CAD. The improvement was statistically significant (p=0.008, 0.010, 0.007, and 0.005, respectively) for each of the confidence rating scales. The partial area index for the four confidence rating scales without CAD was 0.198, 0.204, 0.200, and 0.206, respectively. With CAD, the partial area index was improved to 0.369, 0.365, 0.369, and 0.366, respectively. The improvement was also statistically significant for all four confidence rating scales (p=0.005, 0.004, 0.003, and 0.005, respectively). The average ROC curves for the 10 observers when reading with and without CAD were plotted in Fig. 3 based on the ratings using the original quasi-continuous 100-point rating scale. The difference between the average ROC curves based on the results from the quasi-continuous rating scale and the three discrete confidence rating scales was very small, resulting in overlapping ROC curves. We therefore did not plot the ROC curves for the three discrete confidence rating scales. In Fig. 4, the individual A_z values for the reading conditions without CAD by the 10 radiologists with quasi-continuous 100-point scale and the discrete 5-point scale are compared. A small difference in the A_z values can be observed for the two different rating scales. Similarly, in Fig. 5 the individual A_z values for the reading with CAD for both scales are shown. The difference between the A_z values for the two confidence rating is also small as in the case of reading without CAD mode. In both reading modes (Fig. 4 and Fig. 5) we can observe a little effect on the A_z values when we use different confidence ratings scales.

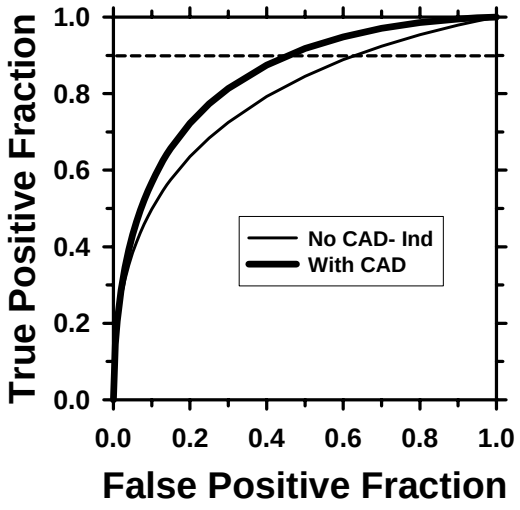


Figure 3. The average ROC curves for the 10 observers when reading with and without CAD.

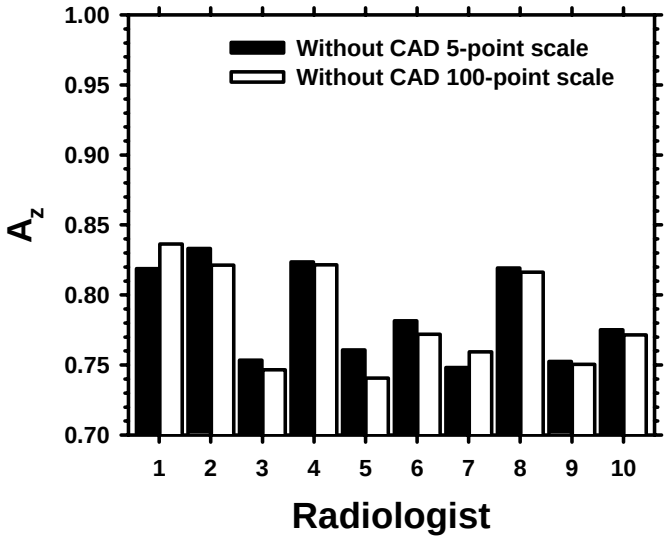


Figure 4. The individual A_z values for the 10 radiologists under the reading condition of without CAD using the quasi-continuous 100-point scale and the discrete 5-point scale.

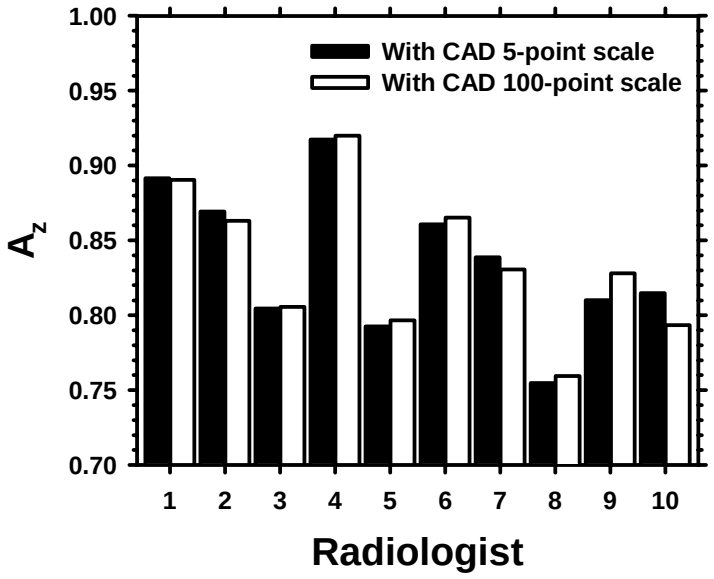


Figure 5. The individual A_z values for the 10 radiologists under the reading condition of with CAD using the quasi-continuous 100-point scale and the discrete 5-point scale.

4. CONCLUSION

We studied the effects of the number of confidence ratings used in an observer experiment on the results of ROC analysis. An observer ROC study that was performed to evaluate the effects of computer-aided diagnosis on radiologists' characterization of masses on serial mammograms was used an example. The original observer ratings were collected on a quasi-continuous 100-point scale. Discrete rating scales of 20, 10, and 5 categories were simulated by grouping adjacent ratings in groups of 5, 10, and 20, respectively. We found that the use of continuous and discrete confidence rating scales had minimal effects on the ROC analysis in this study. The use of CAD significantly improved radiologists' accuracy in classification of masses on serial mammograms for all confidence rating scales examined.

ACKNOWLEDGMENTS

This work was supported by USAMRMC grants DAMD17-02-1-0489 and DAMD17-02-1-0214, and USPHS grant CA95153. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

REFERENCES

1. H. E. Rockette, D. Gur, and C. E. Metz, "The use of continuous and discrete confidence judgements in receiver operating characteristic studies of diagnostic imaging techniques.," *Investigative Radiology* **27**, 169-172, 1992.

2. J. L. King, C. A. Britton, D. Gur, H. E. Rockette, and P. L. David, "On the validity of the continuous and discrete confidence rating scales in receiver operating characteristic studies," *Invest Radiol* **28**, 962-963, 1993.

3. R. F. Wagner, S. V. Beiden, and C. E. Metz, "Continuous versus categorical data for ROC analysis: some quantitative considerations," *Academic radiology* **8**, 328-334, 2001.

4. K. S. Berbaum, D. D. Dorfman, E. A. Franken, and R. T. Caldwell, "An empirical comparison of discrete ratings and subjective probability ratings," *Academic Radiology* **9**, 756-763, 2002.

5. L. M. Hadjiiski, H. P. Chan, B. Sahiner, M. A. Helvie, M. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, et al., "ROC study: Effects of computer-aided diagnosis on radiologists' characterization of malignant and benign breast masses in temporal pairs of mammograms," *Proc. SPIE Medical Imaging* **5032**, 94-101, 2003.

6. L. M. Hadjiiski, H. P. Chan, B. Sahiner, M. A. Helvie, M. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, et al., "Improvement of Radiologists' Characterization of Malignant and Benign Breast Masses in Serial Mammograms by Computer-Aided Diagnosis: An ROC Study," *Radiology* **233**, 255-265, 2004.

7. L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. N. Gurcan, "Analysis of Temporal Change of Mammographic Features: Computer-Aided Classification of Malignant and Benign Breast Masses," *Medical Physics* **28**, 2309-2317, 2001.

8. Y. Jiang, C. E. Metz, and R. M. Nishikawa, "A receiver operating characteristic partial area index for highly sensitive diagnostic tests," *Radiology* **201**, 745-750, 1996.

9. D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "ROC rating analysis: Generalization to the population of readers and cases with the jackknife method," *Investigative Radiology* **27**, 723-731, 1992.

Computer-aided detection of microcalcification clusters on full-field digital mammograms: multiscale pyramid enhancement and false positive reduction using an artificial neural network

Jun Ge*, Jun Wei, Lubomir M. Hadjiiski, Berkman Sahiner,
Heang-Ping Chan, Mark A. Helvie, Chuan Zhou
Department of Radiology, University of Michigan, Ann Arbor

ABSTRACT

We are developing a computer-aided detection (CAD) system to detect microcalcification clusters automatically on full field digital mammograms (FFDMs). The CAD system includes five stages: preprocessing, image enhancement and/or box-rim filtering, segmentation of microcalcification candidates, false positive (FP) reduction, and clustering. In this study, we investigated the performance of a nonlinear multiscale Laplacian pyramid enhancement method in comparison with a box-rim filter at the image enhancement stage and the use of a new error metric to improve the efficiency and robustness of the training of a convolution neural network (CNN) at the FP reduction stage of our CAD system. A data set of 96 cases with 200 images was collected at the University of Michigan. This data set contained 215 microcalcification clusters, of which 64 clusters were proven by biopsy to be malignant and 151 were proven to be benign. The data set was separated into two independent data sets. One data set was used to train and validate the CNN in our CAD system. The other data set was used to evaluate the detection performance. For this data set, Laplacian pyramid multiscale enhancement did not improve the performance of the microcalcification detection system in comparison with our box-rim filter previously optimized for digitized screen-film mammograms. With the new error metric, the training of CNN could be accelerated and the classification performance in validation was improved from an A_z value of 0.94 to 0.97 on average. The CNN in combination with rule-based classifiers could reduce FPs with a small tradeoff in sensitivity. By using the free-response receiver operating characteristic (FROC) methodology, it was found that our CAD system can achieve a cluster-based sensitivity of 70%, 80%, and 88% at 0.23, 0.39, and 0.71 FP marks/image, respectively. For case-based performance evaluation, a sensitivity of 80%, 90%, and 98% can be achieved at 0.17, 0.27, and 0.51 FP marks/image, respectively.

Keywords: Computer-aided detection (CAD), Full-field digital mammography (FFDM), Multiscale pyramid enhancement, Artificial neural network

1. INTRODUCTION

Breast Cancer is the most frequently diagnosed cancer and ranks second among cancer deaths in women. An estimated 211,240 new cases of invasive breast cancer and an estimated 40,410 breast cancer death are expected to occur among women in the US during 2005[1]. Studies indicate that the screening/diagnosis and treatment at early stage can improve the survival rate of women with breast cancer[2-4]. Mammography is the most effective method to date for the detection of breast cancers. However, it has been reported that a substantial fraction of breast cancers which are visible upon retrospective analyses of the images are missed initially[5-7]. The use of a computer-aided detection (CAD) system as an objective 'second reader' is considered to be one of the promising approaches that may help radiologists improve the sensitivity of mammography. Studies have shown that CAD can improve radiologists' detection accuracy significantly[8-10].

Mammographic CAD algorithms were developed for digitized screen-film mammograms before the advent of full-field digital mammography (FFDM). FFDM technology has advanced rapidly in the last few years. Several FFDM manufacturers have obtained clearance from FDA for clinical use to date. We have developed a CAD system for the detection of microcalcification clusters on digitized screen-film mammograms in our previous studies[11-13]. We are

* gejun@med.umich.edu, phone: 734-647-8553, CGC B2103, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0904

Medical Imaging 2005: Image Processing, edited by J. Michael Fitzpatrick, Joseph M. Reinhardt, Proc. of SPIE Vol. 5747 (SPIE, Bellingham, WA, 2005)
1605-7422/05/$15 · doi: 10.1117/12.595753

developing a microcalcification cluster CAD system for digital mammograms acquired by FFDM detectors. In this study, we investigated the performance of a nonlinear multiscale Laplacian pyramid enhancement method in comparison with a box-rim filter at the image enhancement stage and the use of a new error function to improve the efficiency and robustness of the training of an artificial neural network at the false positive (FP) reduction stage of our CAD system.

2. MATERIAL AND METHODS

2.1 Materials

The data set we used in this study contained 96 cases with 200 images. Institutional Review Board (IRB) approval was obtained to collect the mammograms in the Department of Radiology at the University of Michigan. The mammograms in this data set were acquired with a GE Senographe 2000D FFDM system. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of $100\ \mu m \times 100\ \mu m$ and 14 bits per pixel. Most of the cases had two mammographic views: the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view, except for 8 cases that had three views. There are 215 microcalcification clusters in the data set, of which 64 clusters were proven by biopsy to be malignant and 151 were proven to be benign. The true locations of the clusters were identified on each image by an experienced radiologist.

2.2 Methods

The design methodology used for detecting microcalcification clusters on digitized mammograms in our previous study[12] was adapted to FFDMs. The CAD system includes five stages: (1) preprocessing, (2) image enhancement and/or box-rim filtering, (3) segmentation of microcalcification candidates, (4) FP reduction using rule-based classifiers and a convolution neural network (CNN), and (5) clustering of microcalcifications. The block diagram of our CAD system is shown in Fig. 1.

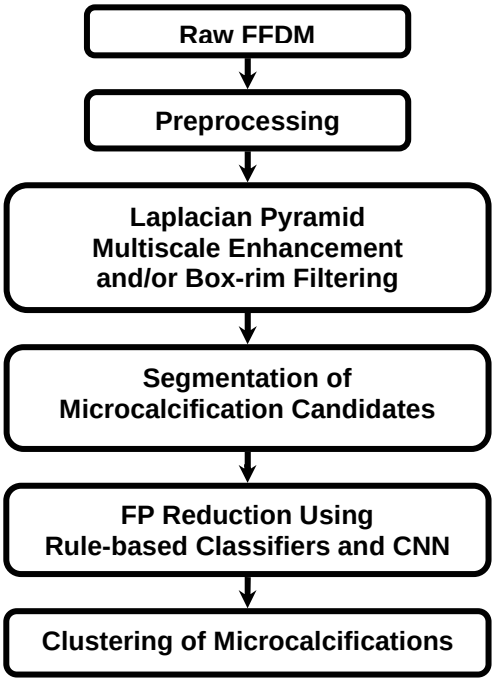


Fig. 1. The block diagram of our CAD system for detection of microcalcification clusters on FFDMs.

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to readers in clinical practice. The image preprocessing method used depends on the manufacturer of

the FFDM system. To develop a CAD system which is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. Clinical mammograms are usually viewed in a negative mode of the raw images. In order to process an image with the same format as the clinical mammograms, we first applied an inverted logarithmic transformation[14] to the raw images in the preprocessing stage. Then the breast boundary is automatically detected and any area external to the breast region is trimmed.

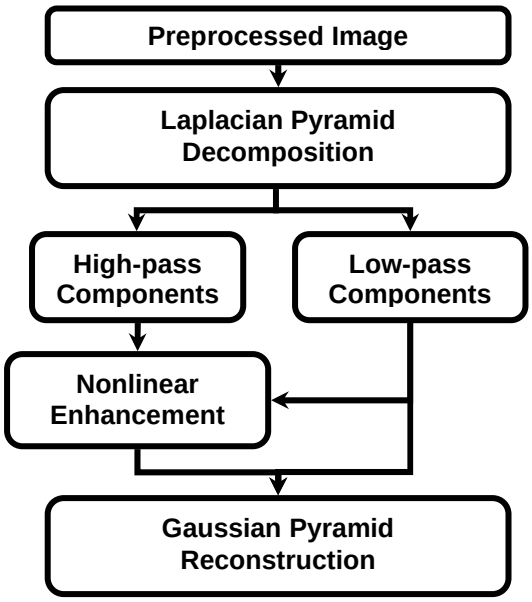
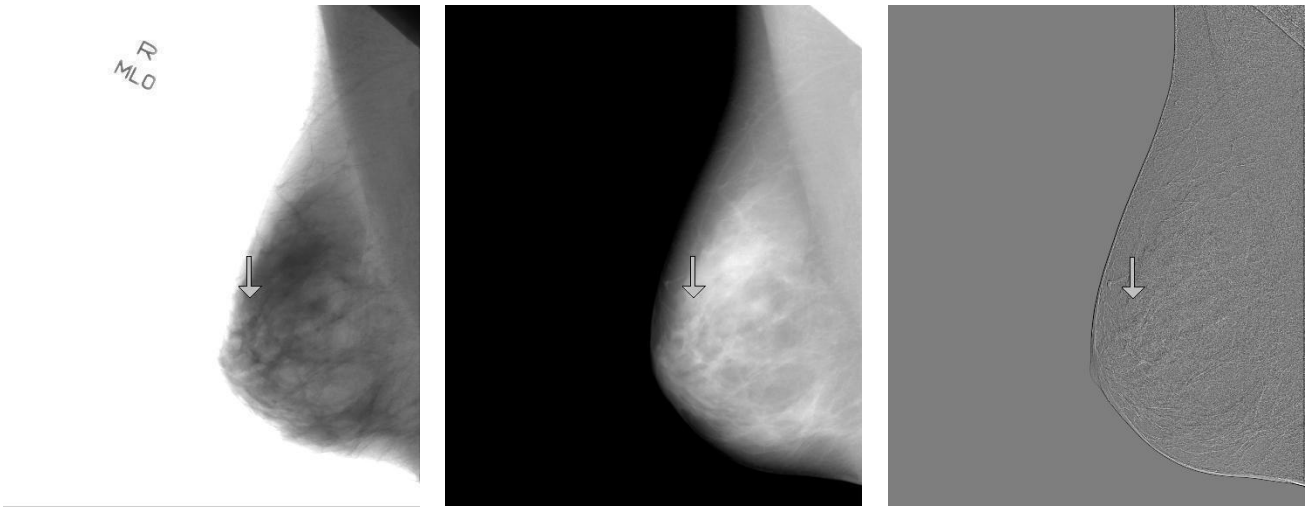


Fig. 2. The schematic diagram for the Laplacian multiscale enhancement method.

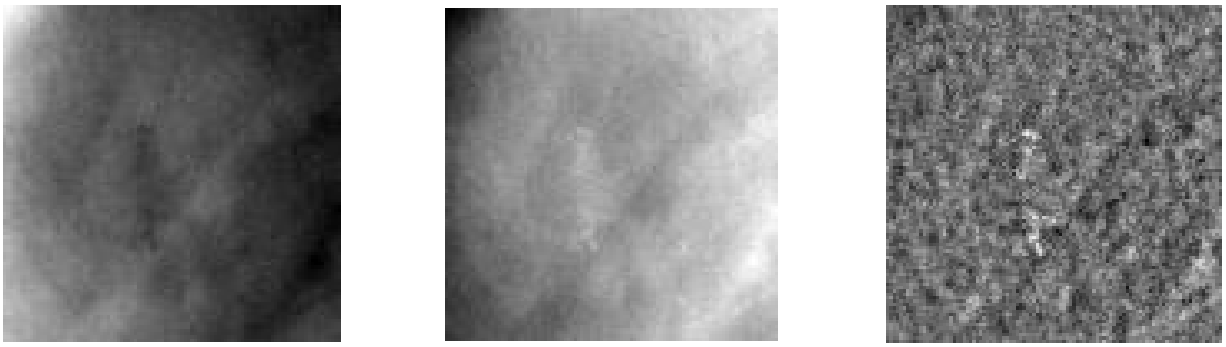
We have designed a multiscale enhancement method for the detection of masses on FFDMs[16]. In this method, the Laplacian pyramid is used to decompose the raw image into multiscale components. A nonlinear weighting function is then employed to enhance the high-pass components and an enhanced image is reconstructed. The nonlinear weighting function used in our study is different from the one used by others[17] in that it explicitly utilized the low-pass band information to aid the enhancement of high-pass components. A Gaussian pyramid interpolation is then used to reconstruct the image from the low-pass components and the enhanced high-pass components. The schematic diagram of the multiscale enhancement is shown in Fig. 2. In this study, we investigated the performance of a nonlinear multiscale Laplacian pyramid enhancement method in comparison with a box-rim filter at the image enhancement stage for the detection of microcalcification clusters. Fig. 3 (a)-(c) show the raw image, Laplacian Pyramid enhanced image, and box-rim filtered image of a mammogram with a subtle microcalcification cluster, respectively. The ROIs containing the true cluster from these images are shown in Fig. 4 (a)-(c), respectively.

In the segmentation stage, potential microcalcification locations are identified with global and local adaptive thresholding methods. The microcalcification candidates after the segmentation stage are shown in Fig. 5(a). In the FP reduction stage, the microcalcification candidates are classified as either true-positive (TP) or FP using the combination of rule-based feature classification and a trained CNN classifier. Two rule-based features used in this study are the area and the gray-scale contrast of the microcalcification candidate. The optimal architecture of CNN has been selected in our previous study[13]. Our CNN was previously trained using back-propagation learning rule with least squares error function. Even with an optimal CNN architecture, the learning curve can oscillate abruptly between iterations when noisy training data are present. In this study, we used a new error function which prohibits the updating of CNN weights when the absolute difference between the CNN output and the target value is larger than a threshold. Since CNN at the first few iterations may deviate far away from the optimal, training samples which produce large error may very well be good data. Thus, the new error function was not applied until after a chosen number of iterations of the training. Finally, clustered microcalcifications are identified by a clustering technique[12]. The TP microcalcifications after this stage are shown in Fig. 5(b). As seen from Fig. 5(b), most of the FP microcalcifications are removed by our rule-based classifiers and CNN.



(a) (b) (c)

Fig. 3. A full-field digital mammogram with a subtle microcalcification cluster. (a) Raw image, (b) Laplacian pyramid enhanced image, (c) Box-rim filtered image of (b).



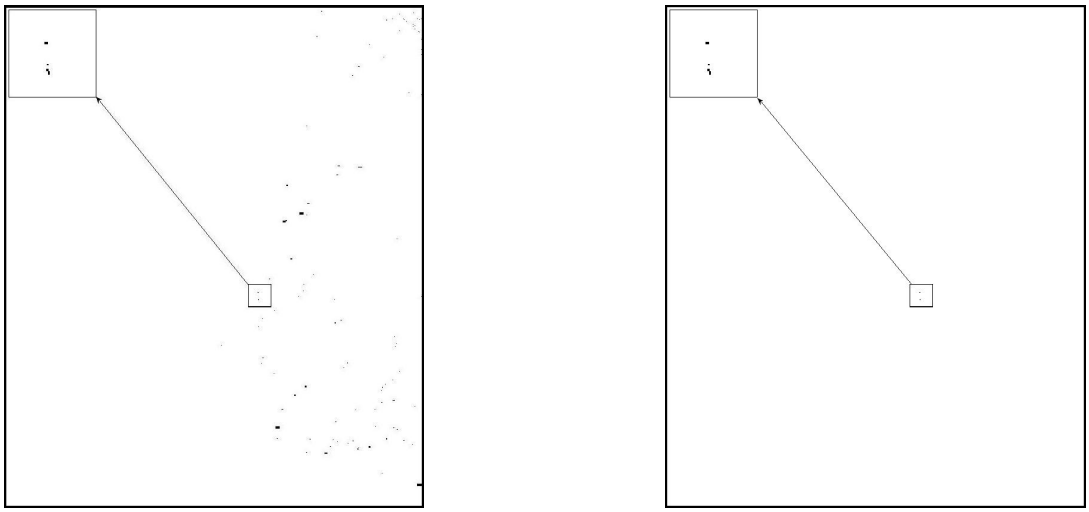
(a) (b) (c)

Fig. 4. The microcalcification cluster in a region-of-interest (ROI) of 1 cm^2 (100 x 100 pixels) on (a) Raw image, (b) Laplacian pyramid enhanced image, (c) Box-rim filtered image of (b).

3. RESULTS

The data set was separated into two independent data sets. One data set was used to train and validate the CNN in our CAD system. The other data set was used as a testing data set to evaluate the detection performance. The testing data set contained 49 cases with 104 images. There are 110 biopsy proven microcalcification clusters in the testing data set. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-cluster and a per-case basis. For cluster-based FROC analysis, the microcalcification cluster on each mammogram was considered an independent true cluster; the sensitivity was thus calculated relative to 110 clusters. For case-based FROC analysis, the same cluster imaged on the two-view mammograms was considered to be one true cluster and the detection of either or both clusters on the two views was considered to be a TP detection. To demonstrate the effects of the Laplacian pyramid enhancement on microcalcification, we disabled FP reduction with CNN for the comparison of the FROC curves with and without enhancement shown in Fig. 6. As can be seen from the FROC curves, for this data set, Laplacian pyramid multiscale enhancement using our currently chosen parameters did not improve the performance of the microcalcification detection system in comparison with our box-rim filter previously optimized for digitized screen-film mammograms. However, it was observed that the

new error metric not only accelerated the CNN training but also improved the classification performance. Our experimental results showed that the performance of the trained CNN was improved from an A_z value of 0.94 to 0.97 on average for the validation set. As a consequence, the number of FP marks/image was reduced as seen from the cluster-based FROC curve in Fig. 7. Under this condition of no Laplacian pyramid enhancement, our CAD system achieved a cluster-based sensitivity of 70%, 80%, and 90% at 2.16, 3.22, and 5.95 FP marks/image, respectively. When CNN was applied, a cluster-based sensitivity of 70%, 80%, and 88% could be achieved at 0.23, 0.39, and 0.71 FP marks/image, respectively. Fig. 7 also shows the case-based FROC curve for our CAD system.



(a) (b)

Fig. 5. (a) Microcalcification candidates after the segmentation stage, (b) Detected microcalcification cluster after the clustering stage.

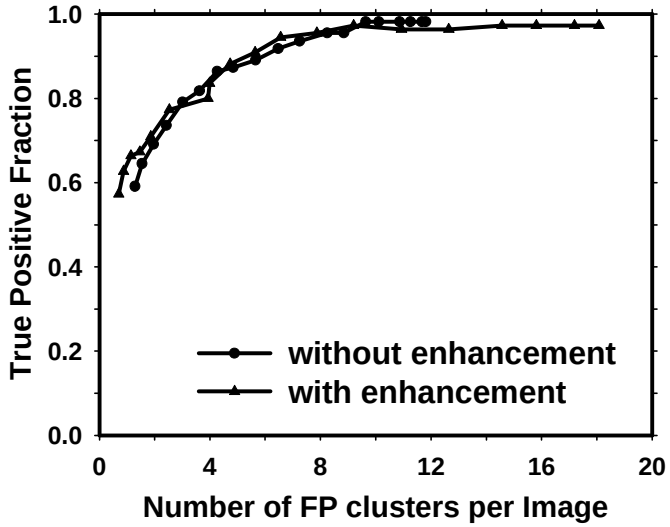


Fig. 6. Cluster-based FROC curves. The FROC curve with dots is the performance of our CAD system without multiscale enhancement. The FROC curve with triangles shows the detection performance on the images with multiscale enhancement. FP reduction with CNN was disabled for both curves.

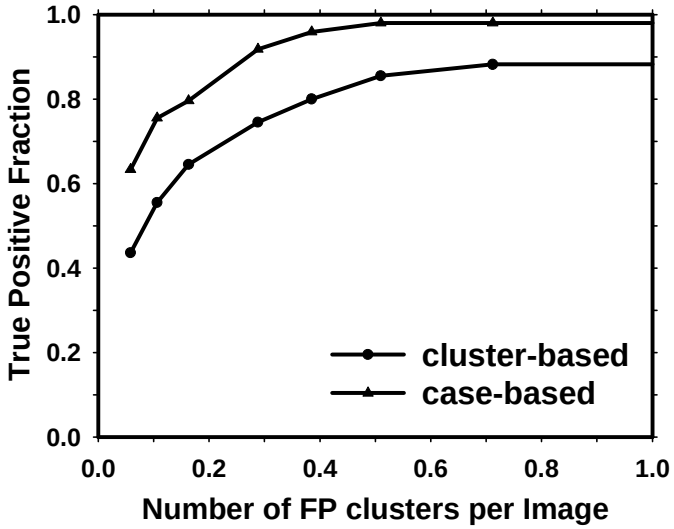


Fig. 7. Overall performance of our CAD system for microcalcification cluster detection. The FROC curve with dots is case-based and the FROC curve with triangles is cluster-based. No multiscale enhancement was used in processing the images for both curves.

4. DISCUSSION AND CONCLUSIONS

In this work, we developed a CAD system for microcalcification clusters which uses the raw FFDMs as the input. The CAD system therefore can easily be adapted to images acquired by FFDM systems from different manufacturers. Our previous CAD system that was developed on digitized screen-film mammograms was adapted to FFDMs. For this data set, we observed that Laplacian pyramid multiscale enhancement did not improve the performance of the microcalcification detection system in comparison with our box-rim filter previously optimized for digitized screen-film mammograms. However, since we have not explored a very wide parameter space for optimization of the enhancement in this study, further work will be needed to examine the effect of image enhancement on the overall detection accuracy. With the new error metric, the training of CNN could be accelerated and the classification performance in validation was improved from an A_z value of 0.94 to 0.97. The CNN in combination with rule-based classifiers can significantly reduce FPs with a small tradeoff in sensitivity. It was found that our CAD system can achieve a cluster-based sensitivity of 70%, 80%, and 88% at 0.23, 0.39, and 0.71 FP marks/image, respectively. For case-based performance evaluation, a sensitivity of 80%, 90%, and 98% can be achieved at 0.17, 0.27, and 0.51 FP marks/image, respectively. Further study is underway to improve the CAD system using a larger data set. In addition, we will incorporate the use of joint two-view information[18] for FP reduction in our CAD system for FFDMs.

ACKNOWLEDGMENTS

This work is supported by USPHS grant CA95153 and U. S. Army Medical Research and Materiel Command grant DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

REFERENCES

1. American cancer society, www.Cancer.Org 2005, "Statistics for 2005"
2. C. Byrne, C. R. Smart, C. Cherk and W. H. Hartmann, "Survival advantage differences by age: Evaluation of the extended follow-up of the breast cancer detection demonstration project," *Cancer* **74**, 301-310, 1994.
3. S. A. Feig and R. E. Hendrick, "Risk, benefit, and controversies in mammographic screening," *Syllabus: A categorical course in physics technical aspects of breast imaging*, A. G. Haus and M. J. Yaffe, Eds., 119-135, Radiological Society of North America, Inc, 1993.
4. C. R. Smart, R. E. Hendrick, J. H. Rutledge and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
5. J. A. Harvey, L. L. Fajardo and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," *American Journal of Roentgenology* **161**, 1167-1172, 1993.
6. R. E. Bird, T. W. Wallace and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613-617, 1992.
7. C. A. Beam, P. M. Layde and D. C. Sullivan, "Variability in the interpretation of screening mammograms by us radiologists - findings from a national sample," *Archives of Internal Medicine* **156**, 209-213, 1996.
8. H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu and H. Macmahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," *Investigative Radiology* **25**, 1102-1110, 1990.
9. L. J. Warren Burhenne, S. A. Wood, C. J. D'orsi, S. A. Feig, D. B. Kopans, K. F. O'shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554-562, 2000.
10. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
11. H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. Macmahon and P. M. Jokich, "Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography," *Medical Physics* **14**, 538-548, 1987.
12. H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Medical Physics* **22**, 1555-1567, 1995.
13. M. N. Gurcan, H. P. Chan, B. Sahiner, L. Hadjiiski, N. Petrick and M. A. Helvie, "Optimal neural network architecture selection: Improvement in computerized detection of microcalcifications," *Academic Radiology* **9**, 420-429, 2002.
14. A. Burgess, "On the noise variance of a digital mammography system," *Medical Physics* **31**, 1987-1995, 2004.
15. C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner and L. M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Medical Physics* **28**, 1056-1069, 2001.
16. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: False positive reduction using gradient field analysis," *Proc. SPIE Medical Imaging* **5370**, 992-998, 2004.
17. M. Stahl, T. Aach, T. Buzug, D. S and U. Neitzel, "Noise-resistant weak-structure enhancement for digital radiography." *Proc. SPIE Medical Imaging* **3661**, 1406-1417, 1999.
18. B. Sahiner, M. N. Gurcan, H. P. Chan, L. M. Hadjiiski, N. Petrick and M. A. Helvie, "The use of joint two-view information for computerized lesion detection on mammograms: Improvement of microcalcification detection accuracy," *Proc. SPIE Medical Imaging* **4684**, 754-761, 2002.

Computer-aided detection of breast masses on mammograms: performance improvement using a dual system

Jun Wei[*a], Berkman Sahiner[a], Lubomir M. Hadjiiski[a], Heang-Ping Chan[a], Mark A. Helvie[a], Marilyn A. Roubidoux[a], Nicholas Petrick[b], Jun Ge[a], Chuan Zhou[a]
[a]Department of Radiology, University of Michigan, Ann Arbor
[b]Center of Devices and Radiological Health, U. S. Food and Drug Administration, Rockville, MD

ABSTRACT

We have developed a computer-aided detection (CAD) system for breast masses on mammograms. In this study, our purpose was to improve the performance of our mass detection system by using a new dual system approach which combines a CAD system optimized with "average" masses with another CAD system optimized with subtle masses. The latter system is trained to provide high sensitivity in detecting subtle masses. For an unknown mammogram, the two systems are used in parallel to detect suspicious objects. A feed-forward backpropagation neural network trained to merge the scores of the two linear discriminant analysis (LDA) classifiers from the two systems makes the final decision in differentiation of true masses from normal tissue. A data set of 86 patients containing 172 mammograms with biopsy-proven masses was partitioned into a training set and an independent test set. This data set is referred to as the average data set. A second data set of 214 prior mammograms was used for training the second CAD system for detection of subtle masses. When the single CAD system trained on the average data set was applied to the test set, the A_z for false positive (FP) classification was 0.81 and the FP rates were 2.1, 1.5 and 1.3 FPs/image at the case-based sensitivities of 95%, 90% and 85%, respectively. With the dual CAD system, the A_z was 0.85 and the FP rates were improved to 1.7, 1.2 and 0.8 FPs/image at the same case-based sensitivities. Our results indicate that the dual CAD system can improve the performance of mass detection on mammograms.

Keywords: computer-aided detection (CAD), mass detection, dual CAD system

1. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 to 55 years of age[1]. It has been reported that early diagnosis and treatment can improve significantly the chance of survival for patients with breast cancer[2-4]. Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially[5-7]. Computer-aided detection (CAD) is considered to be one of the promising approaches that may improve the sensitivity of detecting early breast cancer in screening mammography. It has been shown that CAD can increase the cancer detection rate by radiologists both in the laboratory and in clinical practice[8-13].

We have been developing CAD systems for detection and characterization of mammographic masses and microcalcifications. Detection of masses on mammograms is more challenging than detection of microcalcifications because the normal fibroglandular tissue in the breast causes false positives (FPs) by mimicking masses and causes false negatives due to overlapping with the lesions. Therefore, mass detection systems generally have lower sensitivity and higher FP rate than microcalcification detection systems. In this study, we are investigating the effectiveness of a dual system approach for improving the performance of mass detection on mammograms.

[*] jvwei@umich.edu, phone: 734-647-8553, CGC B2103, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0904

Medical Imaging 2005: Image Processing, edited by J. Michael Fitzpatrick, Joseph M. Reinhardt, Proc. of SPIE Vol. 5747 (SPIE, Bellingham, WA, 2005)
1605-7422/05/$15 · doi: 10.1117/12.595745

2. MATERIALS AND METHODS

2.1 Materials

The data set we used in this study contained 86 cases. Each case included the current mammograms that were obtained before biopsy and the prior mammograms obtained from previous exams. The prior mammograms were used for training the second system because masses on prior mammograms are generally more subtle than those on current mammograms. The subtle mass set does not have to be obtained from the same cases as the average mass set. The current set contained 172 mammograms and the prior set contained 214 mammograms. All data were collected with Institutional Review Board (IRB) approval. The mammograms in this data set were digitized by a Lumiscan laser scanner with a pixel size of $100\mu m \times 100\mu m$ and 12 bits per pixel. All of the current cases had two mammographic views: the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view. There were 86 biopsy-proven masses in this data set. The true locations of the masses were identified by an experienced MQSA radiologist.

2.2 Methods

In order to improve the performance of our CAD system for detection of subtle masses, we developed a new dual system approach which combines a system trained with "average" masses with another system trained with subtle masses. When the trained dual system is applied to an unknown mammogram, the two CAD systems are used in parallel to detect suspicious objects on a single mammogram. No prior mammogram is needed. The additional FPs from the use of two systems are reduced by feature classification in an information fusion stage. Figure 1 shows the block diagram for the dual system.

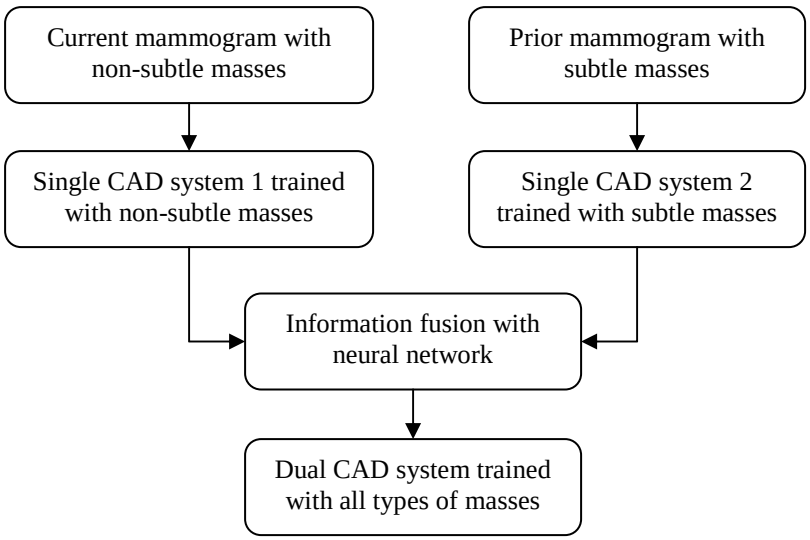


Figure 1. The block diagram of the dual CAD system for mass detection on mammograms.

Our single CAD system consists of five processing steps: 1) digitization, 2) pre-screening of mass candidates, 3) identification of suspicious objects, 4) extraction of feature parameters, and 5) classification between the normal and the abnormal regions by using rule-based and LDA classifiers. The block diagram for the single CAD system is shown in Figure 2. Figure 3 shows an example demonstrating the processing steps with our computer-aided mass

detection system. For the pre-screening stage, we have developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis[14,15]. The gradient field analysis was used to determine locations of high convergence of radial gradient in the image. A region of interest (ROI) of 256×256 pixels is then identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method[16] in which the location of high gradient convergence is used as the starting point. Figures 3(b) and 3(c) show the initial detection locations and the grown objects, respectively. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. The objects whose new gradient convergence is lower than 80% of the original value are rejected. After prescreening, the suspicious objects are identified by using a clustering-based region growing method. For each suspicious object, eleven morphological features are extracted. Rule-based and LDA classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analysis[17,18] are performed in each ROI by using the spatial gray level dependence matrices at different pixel spacings and angular directions. In order to obtain the best feature subset and reduce the dimensionality of the feature space to design a robust classifier, feature selection with stepwise linear discriminant analysis was applied. Finally, LDA classification is used to identify potential breast masses. Figure 3(d) shows the final detected objects, and Figure 3(e) shows the locations of these objects superimposed on the mammogram.

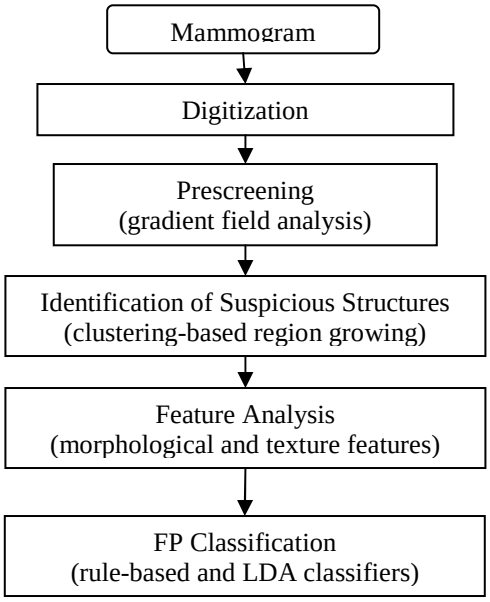
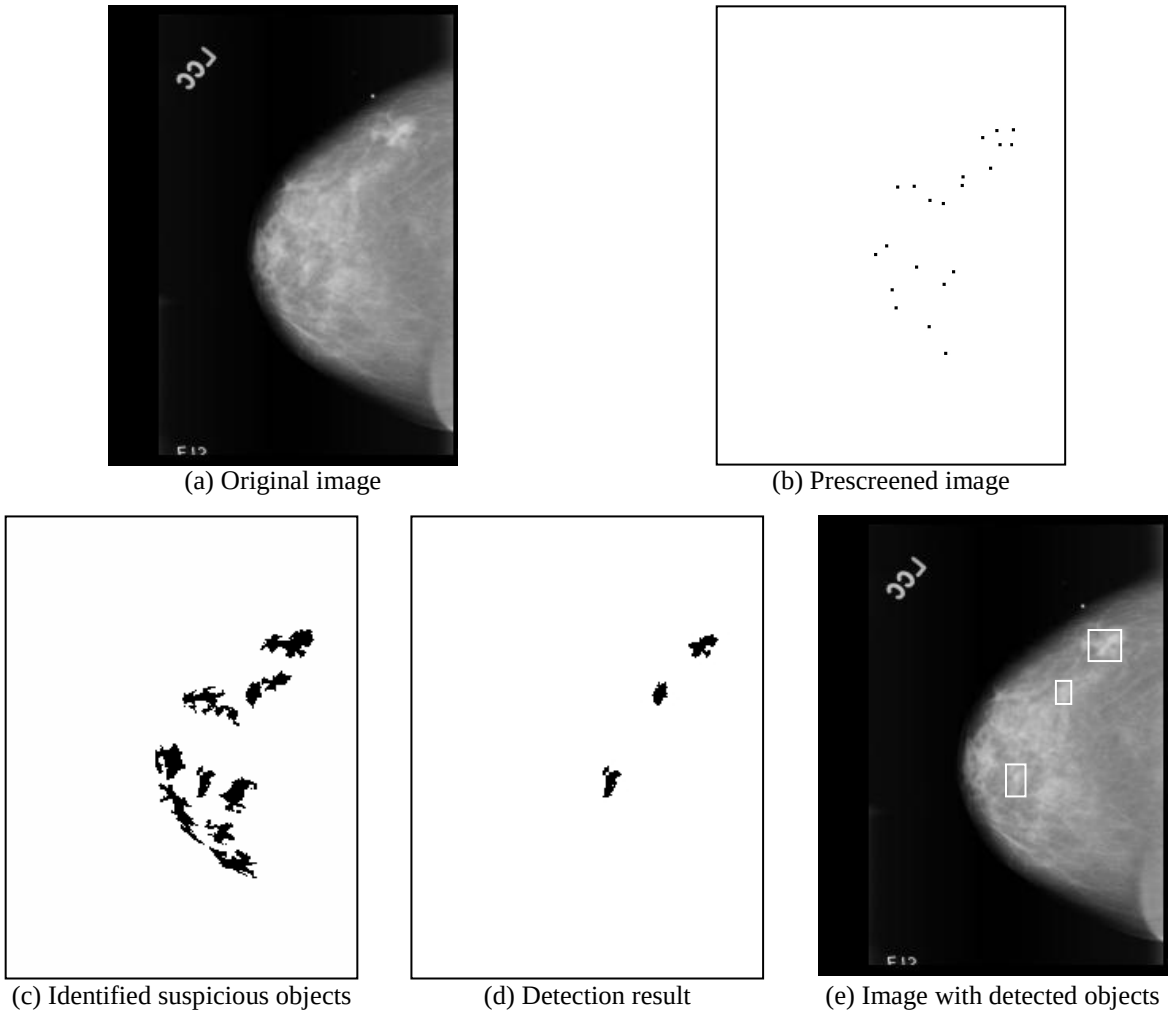


Figure 2. The block diagram of a single CAD system for mass detection on mammograms.

The two single CAD systems were independently trained with the "average" mass set and the subtle mass set, respectively. To merge the information from the two CAD systems, the two LDA discriminant scores from the two CAD systems were used to define a new feature space. A feed-forward backpropagation neural network with 3 hidden nodes was then trained using the LDA feature scores of the training sets as input to differentiate true masses from normal tissue. After the dual CAD system was trained, its performance was evaluated on the independent test set and compared with that of the single CAD system.



(a) Original image (b) Prescreened image

(c) Identified suspicious objects (d) Detection result (e) Image with detected objects

Figure 3. An example demonstrating the processing steps with our single CAD system for mass detection.

3. RESULTS

We randomly separated the cases in our data set into two independent equal sized data sets, each with 43 cases. The training and testing were performed using the cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 86 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP); the sensitivity was thus calculated relative to 43 masses. The average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. When the single CAD system trained on the average data set was applied to the test set, the A_z for FP classification was 0.81 and the FPs/image were 2.1, 1.5 and 1.3 at the case-based sensitivities of 95%, 90% and 85%, respectively. With the dual CAD system, the A_z was 0.85 and the FP rates were improved to 1.7, 1.2 and 0.8 FPs/image at the same case-based

sensitivities. Figure 4 and 5 shows the comparison of the test performance of the single and dual CAD systems by using image-based and case-based average FROC curves, respectively.

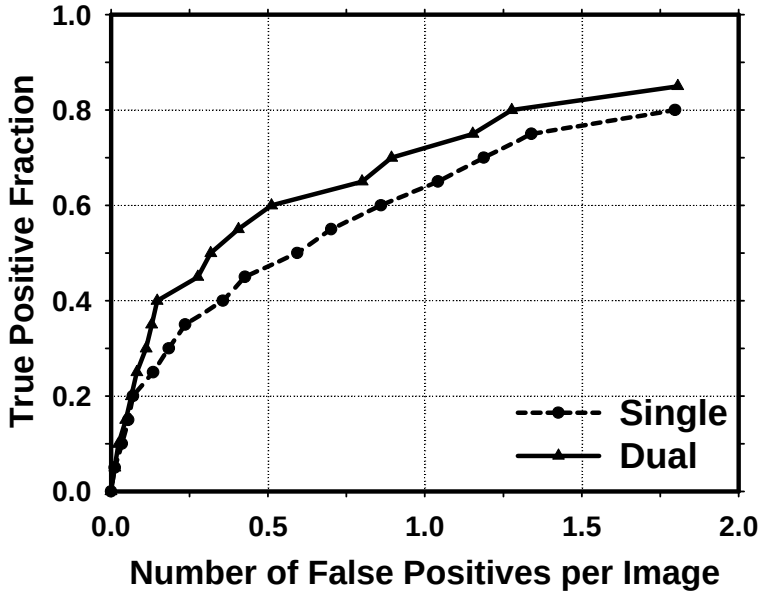


Figure 4. Image-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single: detection by the single CAD system. Dual: detection by the dual CAD system.

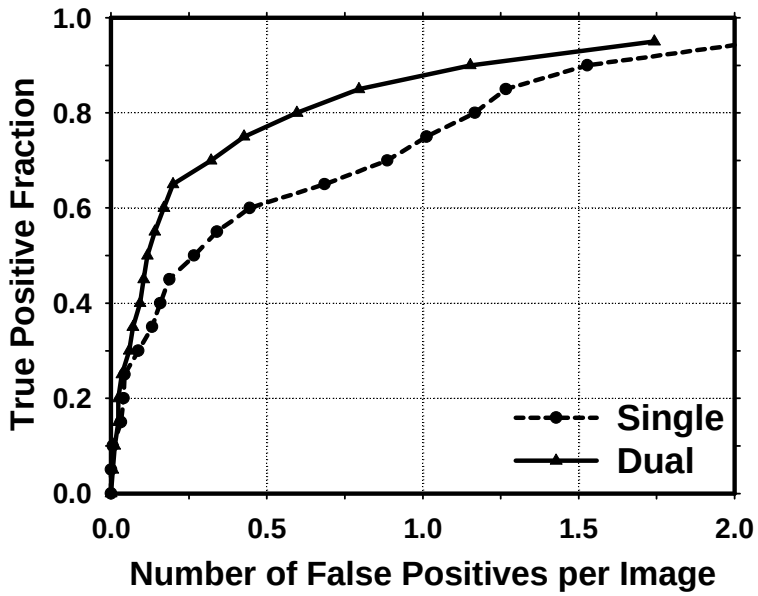


Figure 5. Case-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single: detection by the single CAD system. Dual: detection by the dual CAD system.

4. DISCUSSION AND CONCLUSIONS

We previously developed a CAD system for detection of masses on mammograms. However, we found that it is difficult to train a single system to provide optimal detection for all lesions over the entire spectrum of subtlety. In this study, we developed a dual system which combines a system trained with subtle lesions on prior mammograms and a system trained with masses detected on current mammograms. It was found that the dual CAD system could achieve a higher accuracy than the single CAD system. Further study is underway to optimize the fusion scheme in our dual system.

ACKNOWLEGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD17-03-1-0475 and DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

REFERENCES

1. American cancer society, www.Cancer.Org 2005, "Statistics for 2005"
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," *Am J Roentgenol* **171**, 29-33, 1998.
4. B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *Cancer* **91**, 1699-1703, 2001.
5. C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists - Findings from a national sample," *Archives of Internal Medicine* **156**, 209-213, 1996.
6. R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology* **219**, 192-202, 2001.
7. J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," *J. National Cancer Institute* **95**, 1384-1393, 2003.
8. H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," *Investigative Radiology* **25**, 1102-1110, 1990.
9. L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554-562, 2000.
10. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
11. R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," *Am J Roentgenology* **181**, 687-693, 2003.
12. S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," *Radiology* **232**, 578-584, 2004.

13. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, et al., "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - A pilot clinical trial," *Radiology* **231**, 208-214, 2004.
14. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou, and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis," *Proc. SPIE Medical Imaging* **5370**, 992-998, 2004.
15. J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical.Physics* 2005 (Submitted).
16. N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Medical Physics* **26**, 1642-1654, 1999.
17. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograpms: Multiresolution texture analysis," *Medical Physics* **22**, 1501-1513, 1995.
18. D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis," *Medical Physics* **24**, 903-914, 1997.